Integrative Systems Approaches Towards Brain Pharmacology and Polypharmacology

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Mohammad Shahid

aus

Mardan, Pakistan

Bonn 2013

Angefertigt mit Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der Rheinichen
Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. rer. nat. Martin Hofmann-Apitius

2. Gutachter: Prof. Dr. rer. nat. Holger Fröhlich

Tag der Promotions: 26.02.2014

Erscheinungsjahr: 2014

Abstract

Polypharmacology is considered as the future of drug discovery and emerges as the next paradigm of drug discovery. The traditional drug design is primarily based on a "one target-one drug" paradigm. In polypharmacology, drug molecules always interact with multiple targets, and therefore it imposes new challenges in developing and designing new and effective drugs that are less toxic by eliminating the unexpected drug-target interactions. Although still in its infancy, the use of polypharmacology ideas appears to already have a remarkable impact on modern drug development.

The current thesis is a detailed study on various pharmacology approaches at systems level to understand polypharmacology in complex brain and neurodegnerative disorders. The research work in this thesis focuses on the design and construction of a dedicated knowledge base for human brain pharmacology. This pharmacology knowledge base, referred to as the Human Brain Pharmacome (HBP) is a unique and comprehensive resource that aggregates data and knowledge around current drug treatments that are available for major brain and neurodegenerative disorders. The HBP knowledge base provides data at a single place for building models and supporting hypotheses. The HBP also incorporates new data obtained from similarity computations over drugs and proteins structures, which was analyzed from various aspects including network pharmacology and application of *in-silico* computational methods for the discovery of novel multi-target drug candidates.

Computational tools and machine learning models were developed to characterize protein targets for their polypharmacological profiles and to distinguish indications specific or target specific drugs from other drugs. Systems pharmacology approaches towards drug property predictions provided a highly enriched compound library that was virtually screened against an array of network pharmacology based derived protein targets by combined docking and molecular dynamics simulation workflows. The developed approaches in this work resulted in the identification of novel multi-target drug candidates that are backed up by existing experimental knowledge, and propose repositioning of existing drugs, that are undergoing further experimental validations.

Acknowledgements

First and foremost I would like to express my deep gratitude to my research supervisor, Prof. Dr. Martin Hofmann-Apitius, for providing the opportunity to conduct my research in the department of Bioinformatics at Fraunhofer Institute SCAI. I am very much thankful to him for his continuous support, enthusiastic encouragements and valuable feedback on my research work. Furthermore, I would also like to thank Prof. Dr. Holger Frölich for his willingness to be the co-supervisor of my thesis. My grateful thanks to Dr. Karl N. Kirschner for his guidance and critical feedback during the research work. I would like to extend my thanks to Prof. Dr. Michael Gütschow and Prof. Dr. Evi Kostenis for being part of the PhD evaluation commission.

I would like to extend my gratitude to Wolfgang Ziegler and Horst Schwichtenberg for providing important technical support whenever needed. I also wish to thank my colleagues at Fraunhofer SCAI and B-IT Institute for useful discussions and creating a co-operative working environment.

Publications

1. **Shahid M**, Kirschner K, Hofmann-Apitius M. Pharmacome Tools: A Cytoscape Plugin to Filter Large Drug-Protein Networks for Predicting New Drug-Protein Associations. Submitted to Bioinformatics, Sep. 2013.

2. **Shahid M**, Cheema MS, Klenner A, Younesi E, Hofmann-Apitius M. SVM Based Descriptor Selection and Classification of Neurodegenerative Disease Drugs for Pharmacological Modeling. Molecular Informatics. 2013. 32(3): pp 241-249. DOI: 10.1002/minf.201200116.

3. **Shahid M**, Kasam V, Hofmann-Apitius M. An Improved Weighted-Residue Profile Based Method of Using ProteinLigand Interaction Information in Increasing Hits Selection from Virtual Screening:A Study on Virtual Screening of Human GPCR A2A Receptor Antagonists. Molecular Informatics. 2010. 11: pp 781-791. DOI: 10.1002/minf.201000068.

4. **Shahid M**, Hofmann-Apitius M, Wäldrich O, Ziegler W. A robust framework for rapid deployment of a virtual screening laboratory. Stud Health Technol Inform. 2009;147: pp 212-21.

5. Wolf A, **Shahid M**, Kasam V, Ziegler W, Hofmann-Apitius M. In silico drug discovery approaches on grid computing infrastructures. Curr Clin Pharmacol. 2010. 5(1): pp 37-46.

6. **Shahid M** , Ziegler W, Kasam V, Zimmermann M, Hofmann-Apitius M. Virtual high throughput screening (vHTS) on an optical high speed testbed.Stud Health Technol Inform. 2008. 138: pp 125-34.

7. Ivchenko O, Younesi E, **Shahid M**, Wolf A, Müller B, Hofmann-Apitius M. PLIO: an ontology for formal description of protein-ligand interactions. Bioinformatics. 2011. 27(12): pp 1684-90. DOI: 10.1093/bioinformatics/btr256.

8. Saeed A, Zaib S, Pervez A, Mumtaz A, **Shahid M**, Iqbal J. Synthesis, molecular docking studies, and in vitro screening of sulfanilamide-thiourea hybrids as antimicrobial and urease inhibitors. Medicinal Chemistry Research, 2012. pp 1-10. DOI: 10.1007/s00044-012-0376-4.

9. Ali S, Saeed A, Abbas N, **Shahid M**, Bolte M, Iqbal J. Design, synthesis and molecular modelling of novel methyl[4-oxo-2-(aroylimino)-3-(substituted

phenyl)thiazolidin-5-ylidene]acetates as potent and selective aldose reductase inhibitors. Med. Chem. Commun., 2012. 3: pp 1428-1434. DOI: 10.1039/C2MD20228J.

10. Aslam MA, Mahmood S, **Shahid M** , Saeed A, Iqbal J. Synthesis, biological assay in vitro and molecular docking studies of new Schiff base derivatives as potential urease inhibitors. Eur J Med Chem. 2011. 46(11): pp 5473-5479. DOI: 10.1016/j.ejmech.2011.09.009.

11. **Shahid M**, Klatt T, Rasheed H, Wäldrich O, Ziegler W. SCAI-VHTS - A Fully Automated Virtual High Throughput Screening Framework. ERCIM News. 01/2010; 2010:23-24.

12. **Shahid M**, Alam I, Fuellen G. Biotool2Web: creating simple Web interfaces for bioinformatics applications. Applied Bioinformatics. 2006. 5(1): pp 63-6.

13. **Shahid M** , Mohammad F, Tahir M. Path Coefficient Analysis in wheat. Sarhad Journal of Agriculture. 01/2002. 18(4).

Contents

List of Figures

List of Tables

Dedicated to my beloved father,

dearest Dil Rosh Khan...
gone but never forgotton....

Chapter 1

Introduction and Motivation

1.1 Human Brain Disorders

Brain disorders cause up to 25% of global death and disability [1] and account for more than 34% of the global burden of diseases [2]. Therefore, there is a need for increasing awareness among the community for developing preventive strategies and remedies for these disorders [3]. The social and economic burden arising from brain diseases is even larger than reported because many brain disorders are not accounted for since they arise from factors outside of the nervous system [4], such as those diseases arising from HIV infections [5]. Brain disorders impose high economic burden on society with larger annual costs than the cost of all other disease areas [6]. The significant economic burden on the healthcare system [7] directly and indirectly [8] imposes additional severe impacts on social behavior or leading to suicides [9],[10].

1.2 Neurodegenerative Disorders

With approximately 100 billion neurons and approximately 100 trillion synapses, the complexity of the human brain and its continuous operation renders it vulnerable to abnormal functions and injuries [11]. Abnormalities in specific brain functions, such as memory, movement and cognition, result in neurodegenerative disorders that are characterized by a slow progressive loss of neurons [12]. The loss of neurons in normal physiological aging is not as significant as compared to the loss from neuronal atrophy in selected areas of the brain due to neurodegenerative diseases [13]. There are many known and unknown factors causing the nervous system degeneration that results in the primary connectivity failure in the neuronal circuitry [14], which consequently affects the normal information processing

by the brain and hence, lead to neurodegenerative disorders.
Alzheimer's disease, the most common type of dementia that affects millions of people, has no treatment available today that can reverse or stop its progression [15]. It is clear that Alzheimer's disease and dementia are associated with significant mortality [16]. Parkinson's disease is another age-related common neurological disorder with significant morbidity and disability [17], whose current treatment strategies involve symptomatic management only [18], [19]. The increasing number of available treatment options further complicates the management of the disease [20]. Multiple sclerosis and amyotrophic lateral sclerosis cause neurological disability and shortened life spans in young adults [21]. The limited and partially effective treatment options puts pressure for the development of new and more effective treatments [22] that have reduced side effects[23]. Current treatments for the vast majority of neurodegenerative disorders are only symptomatic, temporary and do not alter disease progression and pursuit of new therapeutic strategies collectively by pharmaceutical industries, academic researchers and disease foundations is the only means to make progress against the challenges presented by these diseases [24].

1.3 Complexities in Disease Mechanims

Neurodegeneration is the principal pathology associated with most neurodegenerative disorders, such as Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis and Huntington's disease [25]. However, controversies in classifying different forms of neurodegenerative disorders often result in unclear demarkation between them [26]. Some disorders (e.g. multiple sclerosis, epilepsy, schizophrenia and tumors) are not usually classified as degenerative while for other disorders there are problems in diagnostically labeling them with correct terminology (e.g. differentiating between Parkinson's and Lewy body dementia, between Alzheimer's and fronto-temporal, and between the dysphasic versions of dementia) [27]. Current classification problems result from multiple pathophysiological processes that are common to different diseases [28]. Even though research on neurodegenerative diseases has progressed in the last four decades and there are journals entirely devoted to the topic [29], the challenges for developing treatments continue to grow [30]. It remains unclear if a single molecular target or a single pathways is involved in the disease progression, and if targeting it would be sufficient to reverse the disease's progress [24].
To understand the multifactorial and complex cellular mechanisms behind neurodegenerative diseases, system biology approaches that include mathematical

models followed by experimental validations are being used by researchers [31]. For example, to predict the conformational state of the mutant protein huntingtin (Htt) that interacts with high number of glutamine amino acids and forming aggregates that are associated with Huntington's disease. *In-silico* systems approaches and modeling will improve our understanding of the complex nature of these "system biology diseases" or "complex diseases" [32]. These complex diseases have multifactorial etiologies and the associated dysfunctions resulting from a large number of interacting biomolecules can be differentiated between a normal and a diseased state when studied at the network level [33]. New therapeutic strategies with multi-functional drugs for various central nervous system targets may enhance the treatment of neurodegenerative disorders [34],[35]. Recent advances in omics technologies and the availability of the resulting data in public databases opens up new areas of research such as systems biology and network pharmacology approaches. Such integrative network approaches will allow us to develop an understanding of the biological context within which molecular level interactions result in adverse phenotypes at the organ and organism level [36].

1.4 The Human Brain Pharmacome

The Human Brain Pharmacome (HBP) is a knowledge base that represents our knowledge on the human brain pharmacology. The HBP pharmacome has been designed and generated as a knowledge resource over the course of this research project. The HBP pharmacome is a first of its kind geared towards gathering and integrating biochemical knowledge about all known drugs and developmental/experimental compounds that are relevant to the human brain disorders. It integrates all the interacting targets of these drugs, pathways involving these targets and phenotypic data associated to these targets. The HBP also includes chemical properties and similarities information calculated for the drugs and proteins structures. To the best of our knowledge, the HBP pharmacome is a unique attempt to retrieve and collect therapeutic data around various human brain diseases and extend it by knowledge concerning pocket geometries, pathways and structural similarities between therapeutic compounds . The HBP will serves its purpose in building a knowledge base that would facilitate biomedical research in the area of complex brain and major neurodegenerative disorders (NDD).

1.5 The Grand Vision of the Human Brain Pharmacome - a Knowledge Base representing Human Brain Pharmacology

The grand vision of the HBP is to build a knowledge base to facilitate the study of integrative systems pharmacology approaches towards understanding brain pharmacology and design new therapeutic strategies. The HBP pharmacome continues to develop as a knowledge resource integrating all relevant and multi-scale biological and chemical knowledge from open source databases, biomedical literature, patent documents, electronic health records, clinical trials outcomes and genome wide association studies. The HBP pharmacome will serve a useful purpose in allowing the study of integrative approaches over the collected data and knowledge represented by the pharmacome.

The complete vision of the HBP is outlined in Figure 1.1, which will be followed by the current state of the HBP to describe how far the vision has been accomplished. The foundation layer of the HBP starts with the integration of relevant data from various public databases which will be extended with the information retrieved from biomedical literature, patent documents and electronic health records. This will allow to exploit mining strategies, on top of the data driven approaches, to retrieve further relevant information for integrating with the existing knowledge. There is a large amount of hidden knowledge available as free-text in biomedical literature and valuable information about intellectual property and scientific inventions in patent documents. Information retrieval and mining strategies will be used to retrieve and gather the valuable information from the millions of these documents. The data collection and retrieval will be followed by a curation pipeline. As the curation effort for such a knowledge resource is substantial, an automated curation pipeline needs to be designed. All the relevant information in the HBP pharmacome will not only be collected, but will be properly organized and interlinked. The pharmacome will be modeled in such a way that would utilize mining. This properly organized knowledge model will facilitate an automated update mechanism that is crucial for the incorporation of new and up-to-date knowledge as it becomes available. The entire information gathering strategies should be designed in such a way that periodic updates can be easily applied. The automated update system will also allow for customized alerts and query system for informing researchers of the appearance of new knowledge in patents and biomedical literature. For this purpose, the HBP data layer will be stored in a structured format such as Resource Description Framework (RDF) format.

FIGURE 1.1: The outline envisioning the Human Brain Pharmacome knowledge base design. The HBP knowledge base will allow easy integration of new knowledge to the existing data, make predictions on top of it, run simulations and validate the new findings.

The disease focus of the HBP is on human brain diseases and major neurodegenerative disorders. Pathways dictionaries and disease specific pathways ontology will be used to retrieve pathways mentions from biomedical literature and the retrieved pathways information will highlight the important pathways modulation by drugs treatments. This will enable understanding true mode of action of the drugs at pathway level. Mapping SNP and phenotypic data within the pharmacome will help to identify common genetic factors that influence a disease state [37]. Similarly, new knowledge will be derived from a series of computations to extend further the pharmacome knowledge base.

The HBP knowledgebase spans over a multi-scale of heterogenous data that allows for building models, proposing hypotheses and making new predictions. The prediction layer includes a variety of modeling approaches and tools that can be applied on the pharmacome data. Such modeling approaches include network models and machine learning-based classifier models. The network models can be applied over the networks of drugs and their associated targets. Network-based predictions could be very useful to identify new and interesting associations between drugs and their targets. Drugs and protein targets can be predicted for their polypharmacological profiles for the design of new and effective multi-target drugs. Machine learning-based classifier models can predict the affect profile of new compounds for a disease area or a specific class of drug targets. The classifier models can also have potential applications in drug re-purposing and the prediction of adverse side effects.

The HBP data at a high resolution, ranging from molecular to atomic level, makes it as another useful resource for the development of virtual screening strategies and molecular simulations. The simulation layer will allow the rapid *in-silico* testing of new compound structures that interact with an array of multiple protein targets. The simulation studies include protein-ligand docking calculations, binding pose evaluation and ranking, and molecular dynamics (MD) simulations. The MD simulations are followed by molecular mechanics-based optimizations around the interaction of chemical compounds with their targets. These simulations will help to identify molecular mechanisms linked to desired or adverse drug actions. The HBP will be able to provide the opportunity of deploying brute-force virtual screening workflows built on docking-MD simulation workflows that can be used for the identification of potential novel chemical scaffolds, multi-target drugs and the prediction of secondary targets.

Finally, the predictions made that are based on the application of modeling and simulation methods on the HBP data would need validations. The validations can be quickly performed by *in-silico* methods and later on by *in-vitro* methods. Realizing this vision will require several years to achieve, the following section provides a firm foundation for the HBP continued growth.

1.6 Current State of the Human Brain Pharmacome

The HBP pharmacome, at its current state, is represented as a relational data model. At present, the HBP pharmacome captures information about all available

drugs that are approved, that are in development and experimental compounds that are therapeutically relevant for human brain disorders. For all of the drugs and experimental compounds, their associated targets and the involved pathways are collected. Side effects data for all approved drugs is also retrieved and saved in the pharmacome. For all of the protein targets, human and mouse phenotypic associations are linked when available. New chemical data is externally generated for existing drugs and therapeutic compounds, along with their protein targets. This data is added to the HBP, enlarging its database and possible usage. Structural data for drug molecules are cleaned, energy minimized and stored in the pharmacome. Structural similarities between the drug compounds are computed by external tools and the results are incorporated in the pharmacome. 3D structures for the interacting protein targets for major neurodegenerative disorders are manually selected, their possible pockets are detected, similarities between their pocket geometries are calculated and the results are incorporated in the HBP. Results of classifier models built for discriminating neurodegenerative disease drugs and GPCRs antagonists are also saved in the form of probabilities that indicate the likelihood of compounds being similar to NDD drugs or being similar to antagonist ligands for a specific class of targets. Currently, the "Pharmacome Tools" [38], an analysis tool developed as a Cytoscape plugin specifically for the drug-target networks obtained from the HBP, allows us to filter large networks for highly specific sub-networks and make predictions of new associations between a set of drugs and protein targets. This tool can also filter the network for polypharmacological protein targets, or select a set of highly specific protein targets in the network context for virtual screening.
Keeping the grand vision of the HBP pharmacome in mind, the following list presents the work done within the current HBP pharmacome. Some of the work has been published and some of the results are ready to be published.

- In the predictions layer of the HBP, machine learning-based classifier models were built on the available drug data for NDD disorders and GPCR antagonists. The methodology and initial results have been published in [39] for classifying NDD drugs and interesting findings for GPCR antagonists classifier are ready to publish.

- An application note about "Pharmacome Tools" to filter large drug-protein network and predict new drug-protein associations has been submitted recently [38].

- In the simulation layer, a framework for the deployment of docking-MD workflows has been successfully established [40] which has been used to execute virtual screening experiments.
- A methodology to increase positive hits selection from virtual screening by molecular docking is published in [41].
- New multi-target drugs are identified by executing MD simulations on a set of multiple targets and the results are ready to publish.

1.7 Unique Features of The Human Brain Pharmacome

The HBP pharmacome is a comprehensive knowledge base resource that brings together existing data from a wide variety of public resources and incorporates new data computed from the data after curation. It is unique in that the drug-target data is manually curated for major NDD disorders. Furthermore, an automated curation pipeline and an automated update mechanism will be integrated in the future that will allow efficiently extending the pharmacome by applying mining strategies. Currently, it includes drug target networks that are analyzed with analysis, and subsequent predictions can be made using tools specifically developed for the HBP pharmacome. The HPB pharmacome allows rapid selection, prioritization and analysis of drug-target combinations, providing highly focused data for further *in-silico* pharmacology studies that includes virtual screening by docking, binding pose selection, high-throughput molecular dynamics simulations and molecular mechanics-based free energy calculations. For supporting integrative systems approaches, phenotypic data from phenomic database and pathways from pathways databases are linked to the drug targets in the pharmacome. Data obtained from the brain interactome, an ongoing PhD project, will also be integrated in the future.

1.8 Related Work and Limitations of Current Information Sources

There exist several knowledge base resources that integrate data about human diseases in general or that specialize in a few individual disorders. Moreover, the integrated data in specialized resources is very limited in dimensionality. The

following is a list of some specialized resources with a brief summary about each resource.

- SMIRDB. The Stanley Medical Research Institute online database [42] is a genomic based database that contains fully annotated clinical metadata and gene expression patterns. This database allows for developing an understanding of the genetic effects of human brain disease particularly bipolar, schizophrenia, and depression but not all NDDs. Therefore, the main focus of SMIRDB lies mainly on gene expression data for selected biopolar disorders.

- INDD. The Integrative Neurodegenerative Disease Database [43] is developed at the University of Pennsylvania (Penn) by converting single disease focused databases comprised from individual clinical research groups at Penn to a new integrated database. The database is intended to be used for comparative studies on major NDD disorders. The INDD has been designed to integrate clinical data about Alzheimer's disease, Parkinson's diseases, Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis from the respective clinical databases that exist in different research groups.

- DND. A Database of Neurodegenerative Diseases [44] includes very limited data sets and exemplifies other specialized resources that collect very specific datasets. The current version of the database provides links to 728 abstracts, 203 genes/proteins and 137 drugs.

- NeuoDNet. NeuroDNet [45] is a platform for constructing and analyzing disease specific genetic networks and boolean networks for 12 NDD disorders, generated from signaling molecules and protein-protein interactions. This database is used for understanding the disease mechanism at network level. The disease networks are based on the human genome data with protein coding genes linked to more than 300 genes reported in clinical studies of neurodegenerative diseases.

- Neurocarta. It is a recently developed database resource [46] that consolidates information about genotype to phenotype associations across multiple resources to explore gene-disease associations. Neurocarta focuses mainly on the genetic basis of neurodevelopmental disorders that contain manually curated information on neurodevelopmental disorders. To understand the genetic functions in the context of diseases, Neurocarta aggregates data from multiple disease gene resources.

- Other specialized resources. Some specialized databases, such as AlzGene or PDGene, focus on data subsets that contain highly specific knowledge about certain diseases. These databases are inherently incomplete with regards to representing knowledge about other neurodegenerative diseases. Generic databases such PharmGKB [47], Therapeutic Target Database (TTD) [48] or Comparative Toxicogenomic Database (CTD) [49] that contain general data have the advantage of retrieving data about any particular disease, but there is a need of substantial amount of efforts for curation. A list of some specialized database resources is compiled in Table 1.1.

TABLE 1.1: List of databases that collect very specific datasets, and their indication areas.

Database	**Indication**
AlzGene [50]	Alzeheimer's disease
AlzPathway [51]	Alzheimer's disease
PDGene [52]	Parkinson's disease
MSGene [53]	Multiple Sclerosis
HD Research Crossroads DB [54]	Huntington's disease
ALSoD [55]	Amytrophic Lateral Sclerosis

As seen above, there exist several resources that make the required data available but are disintegrated and not completely overlapping. To address this, the HBP focuses on aggregating and curating the knowledge around all disorders concerning the brain with a sub speciality for major neurodegenerative disorders. The HBP provides the opportunities at a single place to allow for a) efficient retrieval of the focused data, b) construction of computable models and speculation of hypotheses on top of it, and c) the efficient analysis of the underlying data for specific research questions.

1.9 Utilities of the Human Brain Pharmacome

The HBP captures diverse knowledge about human brain pharmacology for the purpose of allowing *in-silico* systems pharmacology approaches towards understanding of the biochemical interactions within the context of brain disorders. It provides multi-scale computable biological and network models comprised of biomolecular interactions in the disease conditions. The data in the HBP is tightly integrated that represents the relevant knowledge for efficient retrieval. The integrated data and information is used to derive knowledge from computations that enables molecular simulations on top of the HBP knowledge base. From our

perspective, we envision a variety of possible integrative systems pharmacology approaches that could be adapted using the HBP pharmacome. From the relevant biological and chemical knowledge it offers a framework to support competitive intelligence and decision making.
Figure 1.2 outlines the different usage scenarios that come from using the HBP in different perspectives. This outline represents some of the major usage areas of the HBP that include, for example, pharmaceutical industries by integrating the HBP pharmacome in an existing semantic framework based on RDF/triple store, chemoinformaticians that can take the advantage of simulations for activity/ADMETOX predictions, and marketing for the integration of patent information in the context of network pharmacology. Studying integrative systems approaches that combines various tools and technologies in the area of systems pharmacology is one of the main objectives presented in this research work. The systems pharmacology approach is built up on the underlying network models presented in the HBP that consists of biomolecular interaction networks. The network and classifier models are used to identify multi-target drugs and design target class specific or disease area specific focused libraries of compounds from large chemical databases. The drug-target networks in the HBP are analyzed with the help of tools develop specifically for the HBP to guide structure-based and ligand-based design that facilitates running compute extensive simulations. The HBP can be efficiently utilized to elucidate the drugs mechanism of actions at molecular and pathway level to facilitate the design of new drug candidates and repositioning of existing drugs.

1.10 Contribution of the Thesis Work

The research contained in this thesis contributes the following major aspects to the HBP:

- Human Brain Pharmacome design and foundation development
  - Database schema design for structured storage of the HBP data
  - Data collection, integration and curation
- Usage of the HBP pharmacome for drug-target networks analyses
- Systems pharmacology approaches towards drugs properties prediction
  - Prediction of NDD from non-NDD drugs

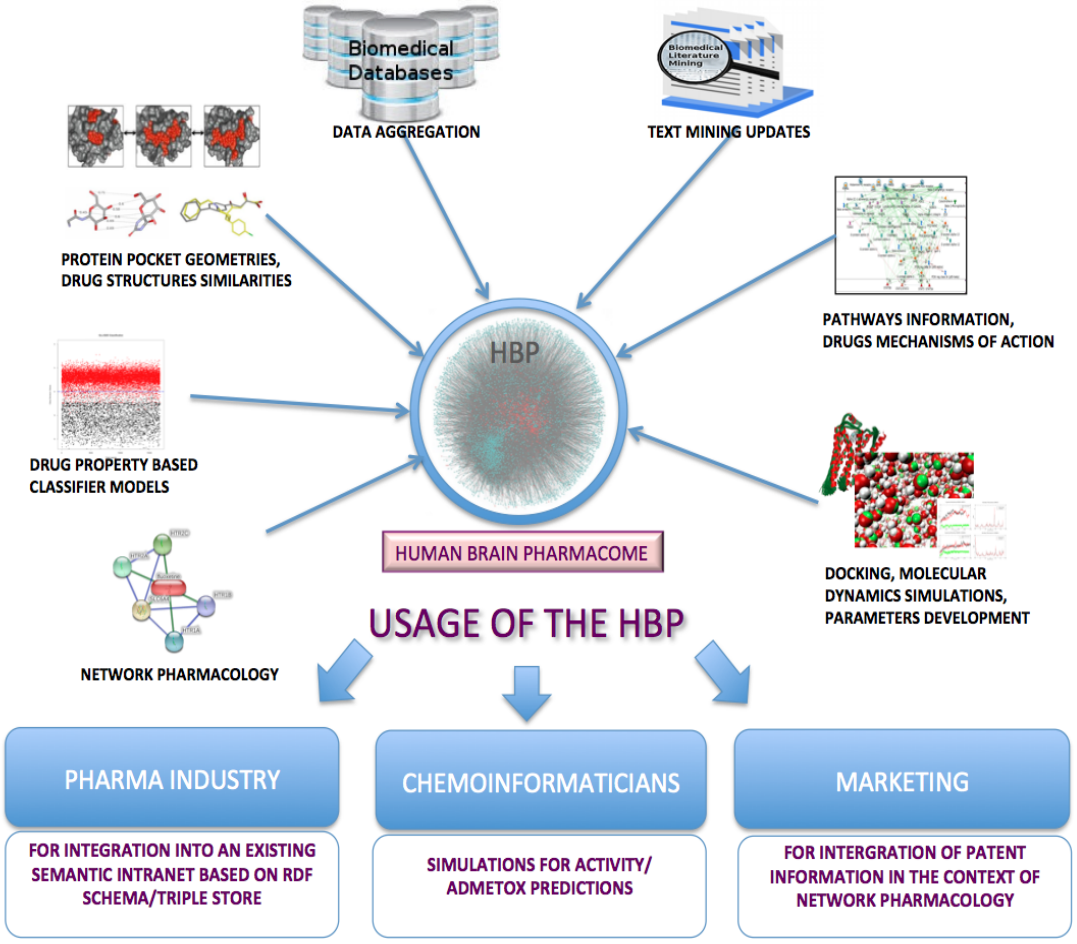


FIGURE 1.2: Integrative systems approaches towards brain pharmacology by utilizing the HBP as a central knowlege base.

  - Prediction of GPCR antagonists from non-GPCR anotagonists

- Approaches towards toxicological targets prediction
- *In-silico* pharmacology approaches towards targeted polypharmacology
  - Exploring polypharmacology by binding site similarity
  - Exploring polypharmacology by chemical similarity
- Running compute intensive simulations to identify novel and re-purposing drug candidates

1.11 Outline of the Thesis

Going along the complete vision of the HBP that is depicted in this chapter, an outline of the integrative systems approaches towards brain pharmacology and polypharmacology as described in various chapters of this thesis, is given below.
Chapter 2 focuses on the HBP data layer and provides an overview of the data sources, the data collection and retrieval process as well as curation and the design of a relational database schema to efficiently store and manage the HBP pharmacome.
Chapter 3 provides applications of network pharmacology within the HBP pharmacome by presenting examples in drug-target networks analyses and network-based predictions. This chapters also gives an overview of the customized pharmacome tools develop for the drug-target networks within the HBP pharmacome.
Chapter 4 and 5 deal with the prediction layer of the HBP from another perspective by providing application scenarios in designing classifier models specific for indication areas or specific target classes. Examples are also provided that have the potential applications in the area of drug re-purposing.
Chapter 6 and 7 represent the simulation layer of the HBP pharmacome. In these chapters, integrative approaches guiding towards both structure-based and ligand-based design methods are exploited to study targeted polypharmacology. The input from the prediction layer is taken forward towards the simulation layer by executing successful deployment of docking-MD workflows to identify multi-target drugs.
The outcome from the predictions and simulations is verified at each step to obtain reliable results. The new findings are taken forward for experimental validations.
Finally, Chapter 8 gives an overall conclusions and future perspectives of our integrative systems approaches towards brain pharmacology.

Chapter 2

Data Collection, Curation and Pharmacome Design

2.1 Introduction

The HBP is a comprehensive knowledge base that collects and represent the knowledge around the human brain disorders and their treatments. It contains a series of chemical and biological networks that represent a scientific understanding of the drugs, their molecular targets, protein-protein relationships, drug-drug relationships, and drug-protein interactions that occur in the brain. Data for the HBP pharmacome is collected from various knowledge base resources that are publicly available. These resources are not focused on specific diseases, instead they offer a specific type of knowledge for many diseases. Currently, there is no single, uniform and standardized resource available that can provide complete and comprehensive knowledge concerning the brain and neurodegenerative disorders at one place. There exist a few specialized genomic databases such as SMRID [42] that provides a comprehensive web-based system for understanding the genetic effects of human brain diseases. Some specialized resources collect genome wide association studies in a single specific disease such as AlzGene [50] for Alzheimer's Disease and PDGene [52] for Parkinson's Disease. Several other attempts have been made to develop integrated neurodegenerative disease databases [44], [43] and several commercial databases that are either focusing on specific topics or are not freely accessible [56]. Another issue is the dispersed pieces of data present in heterogenous resources that are publicly available. In a recent study concerning commercial and public bioactivity databases [57] it was found that each data source contains unique data that is the result of citing different scientific articles. Integrating data from various data sources is useful because there is no complete

overlap in their underlying data. Furthermore, different databases focus on either data subsets or collect a large amount of general data.
This chapter focuses mainly on data aggregation from widely used publicly available database resources. The data is annotated (e.g. addition of structural data) that can be extended later with addition of new knowledge. Manual data curation is performed and a database schema is designed for efficient data storage and handling, that will enable easy retrieval of the required data. Descriptions about all of the individual database resources are provided, that are integrated in the HBP. The major entities of the HBP pharmacome are described. An illustration of the HBP schema design is also presented in this chapter. At the end, an overview of a general workflow is presented that highlights the major components and analysis modules of the research work that leverage on the HBP.

2.2 Data Collection

The data for the HBP as collected from various knowledge resources are described in the following sections. There exist many sources of publicly available knowledge bases that contain information about drug-targets interactions. The knowledge base resources described below are used to extract information about drugs and their targets. These resources are selected on the basis of the triple-store information integrated in their databases, specifically the connected information about a drug, its actual or potential use in a specified disease, and its interacting protein targets. Moreover, various other knowledge base resources are referring to and integrating the information from these databases. A brief description about each knowledge base resource is given below.

2.2.1 MeSH Brain Diseases

A list of all brain disorders, diseases and disease subtypes, which have pathological conditions affecting the brain are present in the MeSH [58] database. These terms can be used as query words to search other publicly available knowledge base resources. The disease list is comprised of more than 270 brain disorders, disease stages and/or subtypes as shown in Table 3.1.

2.2.2 DrugBank

DrugBank [59] is a database of drugs annotated richly with the relevant information concerning the compounds and their targets. Its first version was released in

TABLE 2.1: List of some MeSH disease terms and disease subtypes collected for all disorders in the human brain.

MeSH Disease Term	MeSH Disease Term
Epilepsia Partialis Continua	Alzheimer Disease, Peripheral sensory- neuropathies
Epilepsies, Myoclonic	Alzheimer's Disease
Epilepsies, Partial	Alzheimer's Disease, Dementia
Epilepsy	Amyotrophic Lateral Sclerosis
Epilepsy, Absence	Anxiety
Epilepsy, Benign Neonatal	Anxiety disorder, Chemotherapy-Induced Emesis, Dyspepsia
Epilepsy, Complex Partial	Major- Depressive Disorder
Epilepsy, Generalized	Anxiety Disorders, Depression
Epilepsy, Post-Traumatic	Anxiety or Panic Attacks
Epilepsy, Reflex	Attention-Deficit/Hyperactivity Disorder
Epilepsy, Rolandic	Attention-Deficit/Hyperactivity Disorder, Sleep Disorders
Epilepsy, Temporal Lobe	Multiple Sclerosis
Epilepsy, Tonic-Clonic	Multiple Sclerosis, Chronic Progressive
Epileptic encephalopathy	Multiple Sclerosis, Relapsing-Remitting
Parkinsons Disease	Multiple System Atrophy
Parkinsons Disease 6, Autosomal Recessive Early-Onset	Irritable Bowel Syndrome (IBS)
Parkinson disease, juvenile, autosomal recessive	Schizophrenia, Anxiety and Dementia
Parkinson Disease, Secondary	....
Parkinsonian Disorders	....
Parkinson's Disease Psychosis	....
Restless legs syndrome, Parkinson's disease	....

2006 and has gradually improved with subsequent addition of information. The current version of DrugBank contains approximately 150 data fields for each drug. It is widely used in the field of pharmaceutical research. There are 1424 FDA approved small molecule drugs and 5210 experimental drugs that are interacting with 4326 unique drug targets in version 3.0 of the DrugBank database. This version (accessed in early 2011) was used to extract data for all brain related diseases. The data include the drug names, their interacting biomolecular targets (e.g. proteins or enzymes) and their pharmacological mode of action.

2.2.3 Comparative Toxicogenomic Database

The Comparative Toxicogenomic Database (CTD) [49] is another useful resource of drugs, therapeutic chemicals and target genes involved in human diseases. Its contents include manually curated and literature derived drug-target interactions. Like DrugBank, this database is also cross-referenced by various other knowledge bases that integrate its curated information into their products. In addition to the curated drug-target interactions, a significant number of drug-target associations are derived by inference and provided to the user. Thus, this database also contains inferred associations between chemicals and genes as well as between chemicals and diseases. The CTD database (accessed in early 2011) was used to extract data regarding brain drugs. As of 2011, it contained over 240,300 interactions between 5,931 chemicals and more than 17,000 target genes accounting for more than 3,800 diseases. Only the curated data was extracted and the inference based derived interactions were excluded.

2.2.4 Therapeutic Target Database

As its name suggests, Therapeutic Target Database (TTD) [48] is a useful resource specializing in therapeutic drug targets. The TTD database provides information about therapeutic targets that are mainly proteins and enzymes targets involved in various diseases. It also includes the interacting drugs and their corresponding indications. Both the targets and their drugs are classified as either approved by FDA or those studied in the ongoing clinical trails. TTD database includes 348 FDA approved targets, 292 clinical trail targets and more than 1200 research targets (that are not yet approved therapeutic targets) for all known human diseases. Likewise, there are 1514 FDA approved drugs and 1212 clinical trail drugs that are investigated at various clinical phases I–IV. There are also more than 10,000 preclinical and experimental drugs present in TTD database. The disease annotations are provided separately for both drugs and therapeutic targets in this database. The TTD database (accessed in early 2011) was used to extract data concerning brain drugs and their targets.

2.2.5 EMBL-STITCH Database

The STITCH database [60] from the European Molecular Biology Laboratory (EMBL) is another drug-target interaction network database for small molecules and proteins. The interactions present in the STITCH database are integrated

from more than 20 databases, including DrugBank and CTD, and from the literature. The interactions between small molecules and protein targets are scored based on the experimental evidences, manually annotated data and literature derived information. There are about 500,000 chemicals in total, which are interacting with more than 27,000 human proteins in this database. Interactions present only in human were considered. The interactions are scored in the range from 1 to 999. The higher the score, the higher is the probability that represents the strength of the pharmacological interaction of a chemical with a protein. Version 3.0 of STITCH database (accessed in early 2012) was used to extract the drug-target interactions for all brain drugs There are more than 500 brain related chemicals/drugs in this database that have, on the average, interactions with more than 2000 human proteins.

2.2.6 Activity, Side Effects and Pathways Databases

Apart from the source databases that contain drug-target interactions, there are other specialized databases that were used to annotate the drugs data present in the brain pharmacome. These databases are briefly described in Table 2.2.

2.3 Major Entities and Features Extracted from Source Databases

There are three major entities which are extracted from the above mentioned source databases. These entities are primarily drugs, their targets and the diseases. The drugs and targets are further categorized subsequently into approved, developmental or experimental classes in the context of one or more brain diseases. For each individual entity, all relevant features are extracted that are possibly present in the source databases and additional new features are calculated from external computations. The features or attributes are associated to individual entities that have predefined relationships and are backed up by evidences.

2.4 Data Curation

The pharmacome database is manually curated for the drugs involved in major neurodegenerative disorders (NDD). For each drug, its indication area, its target proteins, its corresponding references to the primary articles or patent documents are checked and merged (if the same drug is present in more than one databases,

TABLE 2.2: List of databases with brief descriptions that were used to extract activities, side-effects and pathways data.

Database Name	Description
ChEMBL [61]	This database contains millions of small drug-like molecules, their 2D structures, calculated properties and bioactivity values. It is maintained by European Bioinformatics Institute (EBI).
BindingDB [62]	This database contains binding affinity data for drug-like molecules against proteins that are considered to be drug targets. BindingDB is an academic project supported by the NIH.
PDSP KiDB [63]	This database contains Ki or affinity values for a large number of drugs and drug candidates for a wide range of targets, including receptors, enzymes, ion channels and transporters.
KEGG [64]	This database resource contains pathways that integrate genomic, chemical and systematic functional information. KEGG pathways are graphical representations of cellular process for the higher order functions of genes.
SIDER [65]	This database is a resource for side effects information of approved drugs. The information about side effects is extracted from public documents and drug labels.

but has additional data or references in the respective database). Similarly, references to the source databases and referenced documents are also checked for correctness concerning individual drugs, targets and drug-target associations. If the indication area is not related to any NDD disorder, it is excluded from the NDD database. When the data is collected from various source databases by using scripts or Java based XML parsers, the results are carefully evaluated for correctness and completeness. It is made sure that data for all FDA approved drugs related to NDD disorders are collected. The redundant data resulting from overlapping entries present in more than one data sources are unified and their corresponding identifiers and references are stored. Similarly, general chemical terms or terms referring to therapeutic classes such as "plan extracts", "xanthine derivatives", "choline-esterase inhibitors" are excluded during the curation, because such terms refer to more than one chemical entities that could not be uniquely identified in the database.

2.5 Pharmacome Schema Design

Two separate schemas are designed for the brain pharmacome to store the data relating to all general brain disorders and NDD disorders (i.e. Alzheimer's disease, Parkinson's disease, Epilepsy and Multiple Sclerosis) in particular. Data concerning all brain disorders are stored in the Human Brain Pharmacome, while the data concerning major NDD disorders are stored in the Neurodegenerative Disorders Pharmacome (NDP).

2.5.1 Human Brain Pharmacome Schema

The HBP pharmacome is stored in a relational database management system running MySQL. The schema for the HBP database (Figure 2.1 is designed for efficient storage and retrieval of the drug-targets interactions and their corresponding attributes or features.

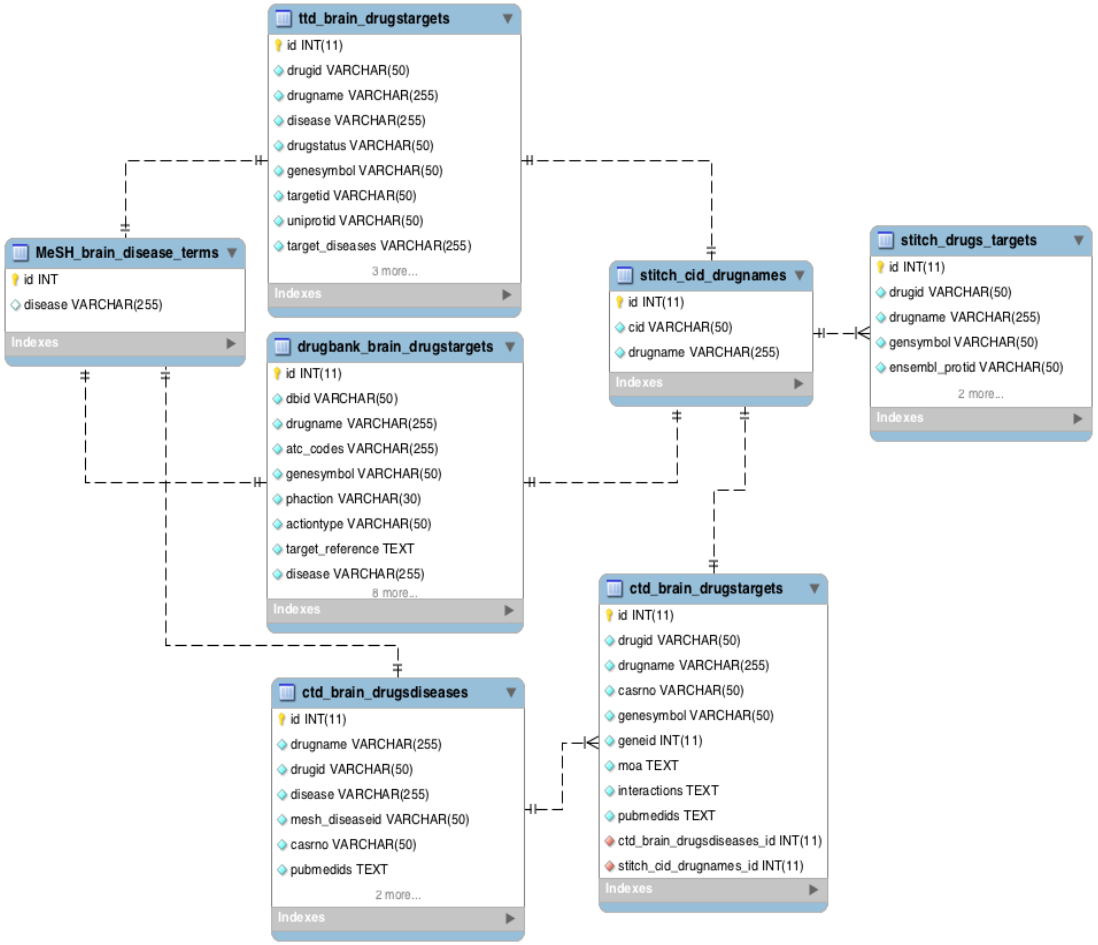


FIGURE 2.1: The HBP pharmacome schema representing relationships between major entities (i.e drugs and their protein-targets that are extracted from various knowledge bases relevant for brain diseases).

2.5.1.1 Major Entity Classes

There are three major entity classes in the HBP, which are briefly described as the following:

1. A MeSH disease term indicating a brain disorder or disease sub-type as shown in Table 3.1 in non-normalized form.

2. A drug or therapeutic compound interacting with a human gene/protein target in a brain disorder indication area. The drugs are referred to by their generic names as the brand/trade names or CAS numbers are not always present for each drug in all of the source databases. Therefore, a drug is uniquely identified by its generic name or the active pharmaceutical ingredient.

3. A gene or its protein product that interacts with a brain drug or compound. An official gene symbol adapted from HUGO [66] gene nomenclature is used to refer to a target gene or their protein products. The unique HGNC gene symbols for the target genes are also mapped to their corresponding UniProt [67] accession number, Entrez gene-id or ENSEMBL protein-id.

For each active compound and target record, the local identifiers are also stored in their respective database tables. Each drug or compounds are distinguished by their drug groups that indicates their current status as an approved or a developmental drug that is in clinical trails. Drugs that are withdrawn from the market or discontinued are separated from the approved and experimental drugs. Similarly, protein targets are also distinguished according to their status as a successful therapeutic targets or as a clinical trail targets. Chemical structures for drugs and compounds, as well as protein crystal structures for target proteins, are also retrieved from the Protein Data Bank (PDB) [68] when they are available. Drugs are annotated with their pharmacological mechanism of action (MOA) or direct interaction type (e.g. increase or decrease of activity) with a protein target. References are also recorded for drugs, target proteins or drug-target interactions when available in the form of Pubmed IDs or patent document IDs.

2.5.2 NDP Pharmacome Schema

A separate database schema is used for storing all drug related data that are involved in four neurodenerative diseases. This schema is similar to the schema of HBP but is manually curated for Alzheimers Disease (AD), Parkinsons Disease

(PD), Epilepsy/Seizures and Multiple Sclerosis (MS). All FDA approved, developmental and experimental drugs, plus therapeutic chemicals, that are used in NDD indication areas are stored in NDP pharmacome, the schema is shown in Figure 2.2. The major entity classes in the NDP pharmacome are the same as described for the HBP pharmacome.

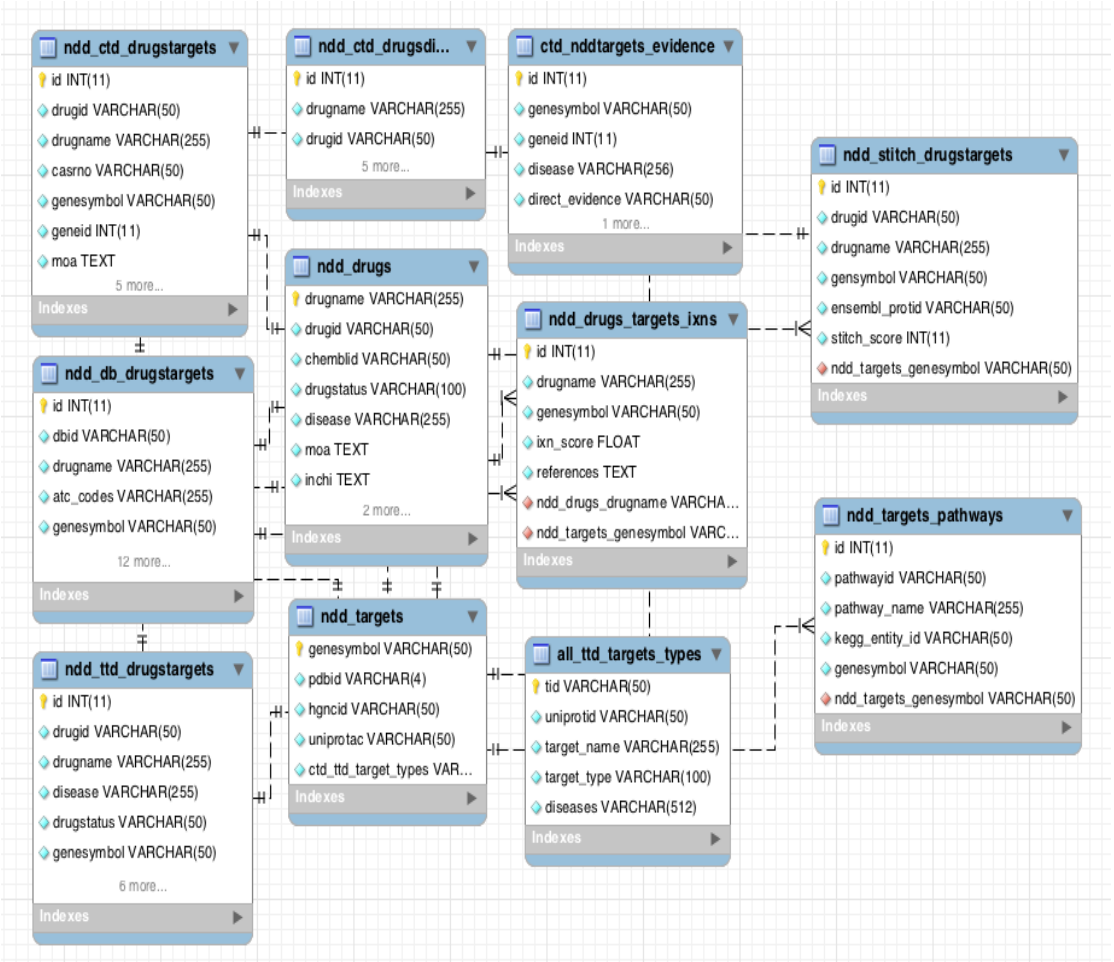


FIGURE 2.2: NDP pharmacome schema that shows the relationships between the drugs and their targets involved in major neurodegenerative disorders.

2.5.2.1 Collection of Compound Structures

Molecular structures of compounds are obtained from DrugBank and TTD databases. If a structure for a compound is not available in these databases, it is searched and obtained from STITCH database. The structures are saved in the form of SMILES strings in the pharmacome.

2.5.2.2 Annotation with Acitivity and Computed Features

The data within the NDP pharmacome is further annotated with activity data, side-effect data and pathway data from specialized databases. Activity data is extracted for all NDD drugs from ChEMBL, KIDB and BindingDB databases. Side effects are retrieved from the SIDER database, and pathways are retrieved from the KEGG database, by focusing on pathways involving drugs and diseases in human (Figure 2.3).

Drug-drug chemical structures similarities and protein-protein binding site similarities are also computed by using external software tools and the results are imported into the NDP pharmacome as drug and target features. Details of drug-drug and protein-protein binding site similarities are presented in the chapter "Drug Target Networks Analysis".

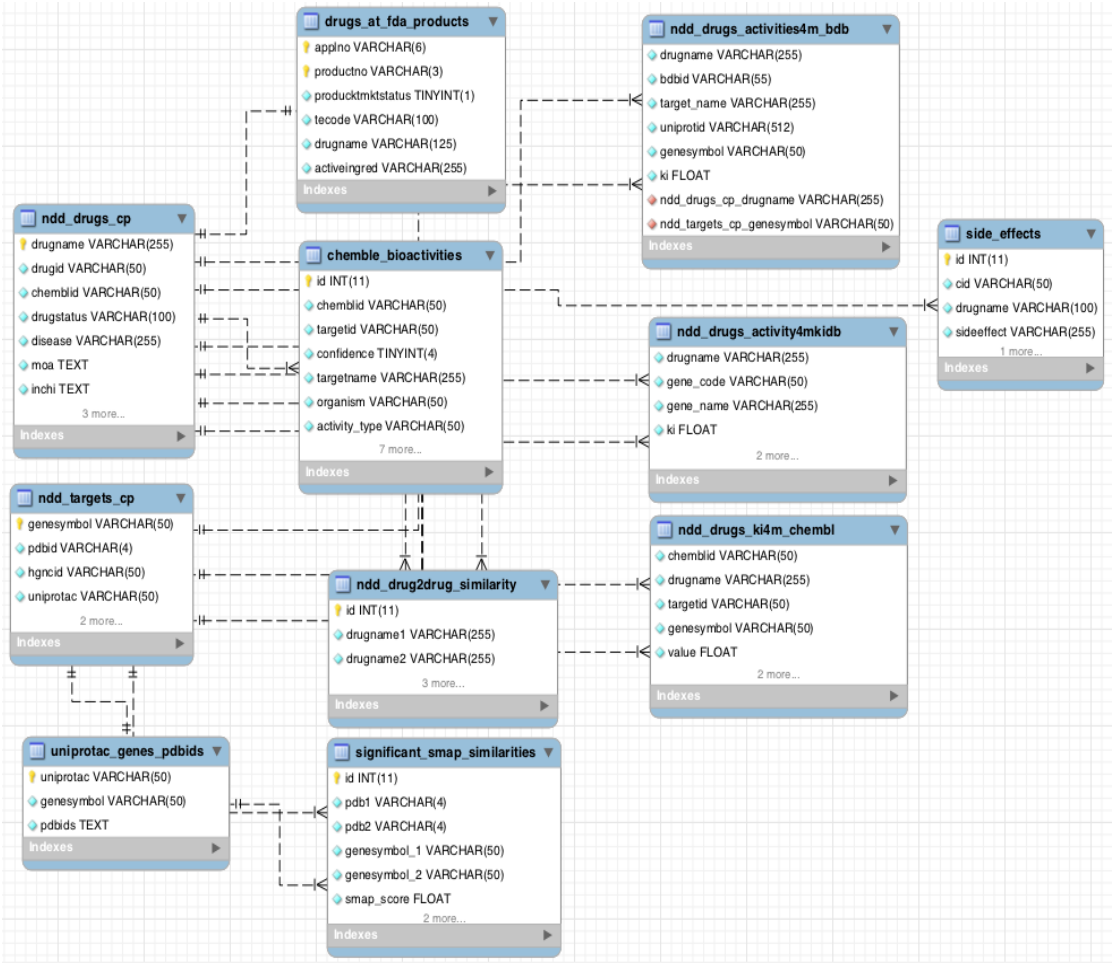


FIGURE 2.3: NDP pharmacome where activity data for drug-target interactions, pathways, side effects and calculated structural similarity values are stored.

2.6 HBP Representation as RDF

Information in the HBP pharmacome is structured in the form of an RDBMS storage and is being converted to RDF format (a common representation format for knowledge that supports its modeling using graphs). RDF format will represent the HBP data in triples, whose components can be traced back to their original sources and will make the addition and updating of new knowledge very easy. The RDF models will allow the easy integration of heterogeneous data present in the HBP pharmacome and provide a framework for future extension and to allow the pharmacome's integration into other existing semantic frameworks.

2.7 Statistics for the Complete HBP

1. >3,500 Drugs and active compounds.
2. >500 Brain disorders and/or disease subtypes.
3. 14,000 Interacting protein targets.

2.8 Statistics for the Complete NDD-focused HBP

1. 700 Drugs (approved drugs, in clinical trials, and discontinued drugs)

   648 Compounds with chemical structures.

2. 5,600 Protein targets for 4 major NDD disorders.

   ∼2,000 Protein targets with 3D structures.

   7,255 Unique genes involved in protein-protein interactions.

   482 Identified successful/therapeutic drug targets.

   196 Identified clinical trial targets.

   487 Identified research targets.

   1,956 Targets with marker and mechanisms in NDDs.

3. ∼23,000 Target pair-wise binding site similarities extracted from ∼2 million pairs of similarity calculations.
4. 3,046 Drug pair-wise chemical similarities.

2.9 Conclusions

Often there is no stand alone database publicly available that contains all the required resources regarding a specific area of interest at a single place. It is therefore, necessary to combine multiple resources to achieve the required data that is as complete and as comprehensive as possible. The extent of data completeness can be obtained by bringing together dispersed pieces of relevant data scattered in multiple heterogenous resources. An obvious advantage of collecting data from heterogenous resources about a common topic of interest is the availability of the data to a possible extent in a predefined and well formed structure enabling efficient reuse of the data later on. A framework for integration of data from multiple data sources is thus very necessary. Therefore, designing an efficient schema for the database plays a vital role, because new and additionally computed data such as chemical similarity and binding site similarity data can be easily integrated. The knowledge represented in the HBP knowledge base spans a multiscale range of resolutions - from systems biology to atomistic information. This multiscale resolution of data will allow us to use the data in several ways that include the guidance of both ligand-based and structure-based design for the rapid qualification of novel chemical entities and reuse of existing drugs in other indications. One of the main goals of developing the HBP knowledge base in this research work, is to have a pharmacology knowledge base that enables the exploration of integrative systems approaches towards brain pharmacology and polypharmacology. An automated update mechanism is very important to frequently update the HBP with new knowledge that arrives in the subject domain.

Chapter 3

Drug Target Networks Analysis

3.1 Introduction

There has been tremendous success in drug discovery and the use of new therapeutic agents during the past few decades [69]. However, in the recent years, the number of new drug approvals has decreased [70]. This lack of success and the decline in new drugs suggests limitations in our understanding of human pharmacology [71]. Moreover, its not always clear why some drugs work successfully while other drugs produce adverse side effects after their extensive use [72]. It is unclear because both the successful ones and those leading to adverse effects have apparently similar properties. To understand the beneficial and adverse effects of drugs, analytical methods such as integrative systems biology and network-based approaches are useful. Using network pharmacology approaches can not only help understanding existing drug-target relationships, but also find new applications for existing drugs [73]. Understanding drug-target networks can also help predict polypharmacological profiles of drugs and predict side effects by discovering off-targets for existing drugs [74], [75].

Network pharmacology is being considered as an important new paradigm in drug discovery [76]. Understanding polypharmacology in the light of integrated network biology enables the expansion of available space of druggable targets [77]. However, there are considerable challenges coming in the rational design of drugs with polypharmacology. For example, how to avoid unwanted polypharmacology [78]. While current methods of uncovering the complex relationships between the drugs and their targets have failed, there is a need for the development of new methods to validate the polypharmacological properties of drugs and their effects [79]. Network pharmacology is one of the new tools that could help develop new chemical entities and treatments by relating the drug-target interactions with patients

phenotypic data drug treatments [80].
Graphs and networks concepts that play important roles in a wide variety of disciplines, are also widely used in the field of chemical and biological informatics [81]. With the availability of high-throughput assays, computational tools, and network analysis through graph theory, its currently possible to analyze large-scale biological networks (e.g. protein-protein interaction networks [82]). The large amount of data that is coming out of genome-wide gene expression profiles, protein-protein and protein-drug interactions generate enormous complexity if their interrelationships are studied [83]. Network-based approaches are used in many steps of the drug-discovery pipeline that help to facilitate visualization, interpretation and analysis of the available heterogeneous data for understanding the molecular basis of diseases [84], [85], [86].
Even with increasing popularity of the network-based approaches in the drug discovery pipeline, there is, however, a limitation on the availability and completeness of the drug-protein and protein-protein interaction data. The interaction data that is available is far from complete and a portion of the available data is biased toward a certain area of interest [87]. In the context of networks, a complete data set would include all experimental data for all biological interactions between protein-protein and protein-drugs when they are physically possible. Inconsistency may arise in networks with incomplete data that would effect the underlying network models. Therefore, care should be taken when deriving conclusions from analyzing such data. In spite of the complex and integrated heterogeneous data collected from various sources, the topology of drug-target interaction networks implicitly depends on data completeness of the drug properties and target families [88].

3.1.1 Targeted Networks Analysis

Adapting systems pharmacology approaches to study the complex associations between the drugs and targets at a network level will not only help understand the underlying complexities, but will also facilitate the discovery of new and interesting relationships. These new insights should help reveal a drug's mechanisms of action and off-target effects, and thus improve drug discovery for complex diseases [89]. To maximize the potential advantages of network-based approaches, it is essential to have a systems-level perspective of a disease. The systems-level perspective that includes interactions and their effects at the molecular-level will help understand the physiological function of the system [90]. Once the network level view representing a complete system is generated by the accumulation of

all experimentally available interaction data for drugs and their targets, the complexity of the network will also increase. Therefore, dealing with large amount of interaction data to identify patterns, the requirements for new analytical and visualization tools become higher [91].
This chapter focuses mainly on the creation, visualization and analysis of a large network describing drug-targets associations involved in neurodegenerative diseases. This network is a part of a larger network entitled as the Human Brain Pharmacome (HBP). By focusing on major NDD disorders, the HBP is filtered to create the specialized NDD pharmacome (NDP). Different sub-networks (e.g drug-drug similarity network, protein-protein pocket similarity network) are generated from the complex NDD network and analyzed for the perspective of polypharmacology. Description about input data preparation, networks creation and analyses are reported. Network-based predictions of secondary targets for existing drugs are made. Network-based clustering of drugs are evaluated for the identification of polypharmacological scaffolds. Functionalities of the Pharmacome tools are described, which is a Cytoscape plugin developed specifically for the analysis of HBP drug-target networks. With the help of Pharmacome tools, targets with polypharmacological profiles are identified and selected for virtual screening.

3.2 Methods and Material

In order to create and analyze drug-target networks, a series of input data preparation steps are performed. The input data is prepared using various tools and methodologies. The following is a list of software tools that are used in the creation and analysis of drug-target networks:

a Molecular Operating Environment (MOE [92]): For molecule 3D structure generation and similarity calculation.

b SMAP software package [93], [94]: For pocket similarity calculation.

c Cytoscape version 2.8.2 [95]: For network visualization.

[i] ClusterMaker plugin for Cytoscape [96], [97]: For clustering drug-target network.

[ii] Pharmacome Tools plugin for Cytoscape: For new interactions prediction, polypharmacological filtering of drugs, protein targets and molecular scaffolds.

3.2.1 Input Data Preparation

Data for drugs and their targets involved in neurodegenerative disorders are retrieved from the HBP database by querying the database for specified attributes (e.g. disease and status attributes for all NDD drugs). The HBP database is described in detail in the chapter Data collection, Curation and Pharmacome Schema Design. The output of the queries is exported to comma or tab-separated formatted files that contain the generic drug names, their associated NDD indications, their current status and all of their interacting protein targets.
For four major NDD disorders, there are approximately 700 compounds that includes approved, clinical trial and discontinued drugs, and more than 5600 protein targets. The network underlying the NDD pharmacome has the following three different classifications for entities:

1. Approved, developmental and discontinued drugs with therapeutic roles in NDDs
2. Disease indication area
3. Disease associated human protein targets encoded by gene symbols

The interactions between drugs and their targets involved in NDD indications lead to a total of 30,211 associations, resulting in an average of ∼40 targets that interact with a single drug. In fact, about half of the NDD drugs interact with 40 to 60 protein targets as shown in Figure 3.1.

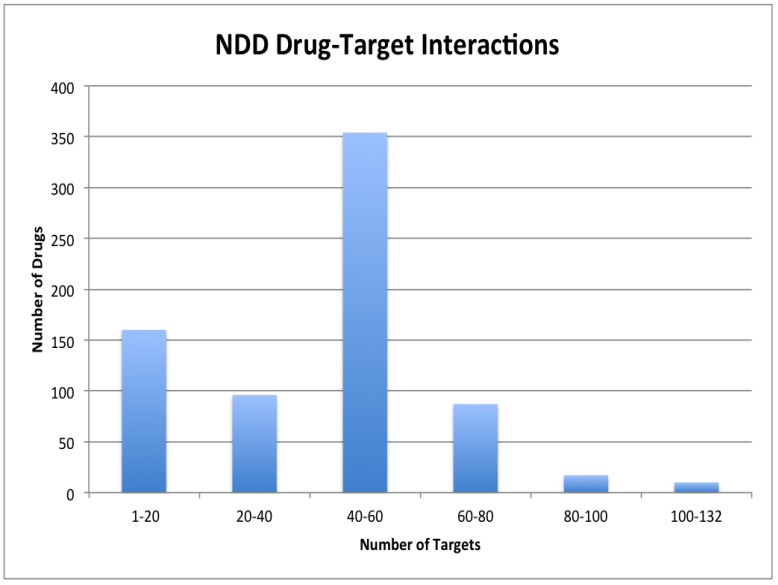


FIGURE 3.1: NDD drug-target interaction statistics for the drugs present in NDD pharmacome.

Drug-target interaction data is prepared initially as a simple interaction format (SIF), which is one of the standard formats required as an input for networks generation. In addition to the drug-target interactions present in the network, the network is extended and populated with externally calculated data. This external data is the result of similarity calculations between drugs chemical structures and similarities of the binding sites (i.e. pockets) of the proteins. The drug-drug similarity and protein-protein binding pocket similarity calculations are described below.

3.2.2 Drug-Drug Similarity

For all NDD drugs, their structural data is also retrieved from the HBP pharmacome database. There are approximately 650 drugs with structural data is available from various databases that include Drugbank [**?**], TTD [48] and STITCH DB [60]. Using the MOE [92] software, the drug structures are prepared by removing salts/counter ions, their 3D coordinates are generated from their 2D SDF files. Neutral protonation species of molecules are generated and finally molecular mechanics based energy minimization with MMFF94x (built-in in MOE) force field is performed. Once the drug structures are prepared, then MACCS structural keys [98] fingerprints are generated and using SVL [92] scripts, pair-wise similarities are computed between all drugs. The resulting similarity matrix is then filtered using a threshold to identify the comounds with highest similarities. Tanimoto-Coeffiecint (TC), a widely used similarity metric [99] is used for setting the similarity cut-off value. Similarities between drugs with TC $\geq$ 0.6 are saved resulting in a total of 3,046 pair-wise similarities between all NDD drugs within the HBP. A higher threshold value is selected in order to reduce the number resulting edges in the similarity network. We did not consider all possible similarity combinations (which are more than 200,000) between the drugs to keep the resulting similarity network smaller and reduce its complexity.

3.2.3 Protein-Protein Binding Pocket Similarity

For all the targets present in the NDD network, that have experimental or theoretical structures available, pair-wise binding site similarities are computed. To compute binding site similarities, structures require some preparation. First, for all protein targets, their unique identifiers in the form of official gene symbols [66] are retrieved from the HBP pharmacome database. The corresponding PDB (Protein Data Bank) [68] identifiers are used to download the 3D structures from the PDB database as mentioned in Chapter 2. There are approximately 2,000 protein

targets in the NDD network with available 3D structures: for these targets, in total, there are more than 9,000 available coordinate files. These structures are filtered in the following semi-automated manner. For a protein with multiple PDB structures (e.g from X-ray, NMR, theoretical methods), the one produced using X-ray crystallography is preferentially picked. If there are multiple X-ray structures, then the structure with best resolution is selected. If there are no X-ray coordinates available, then the first model in NMR structure bundle is selected. Protein structures with multiple chains of the individual structures are manually analyzed with molecular visualization tool Chimera [100]. If the multiple chains are homologous and the ligand-binding site is not located between multiple chains, then only a single chain is retained. Once, the structures are selected, all the co-crystalized bound ligands, cofactors, salts and solvents are removed. The SMAP software [94] package is used to compute binding site similarities between proteins that is based on the sequence order-independent profile–profile alignment (SOIPPA) [93] algorithm. SMAP detects the local binding sites in protein structures and performs similarity comparisons of the protein 3D structural motifs independently of their sequence order. Binding site similarities are computed for more than $\sim$2000 proteins, leading to more than two million pair-wise similarity calculations. Once, the similarity comparisons are completed, the results are filtered for significant similarities. Significant similarity, as defined in the SMAP algorithm, to exist between a pair of proteins when the SMAP score is statistically significant at a given p-value (1.0×10^{-4}) threshold and Tanimoto-Coefficient overlap of the two binding pockets is above a certain threshold (≥ 0.5). From two million comparisons, approximately 23,000 pairs are filtered that have significant similarities.
A defined pocket (e.g. the one occupied by a co–crystallized ligand) of a protein was not used for the similarity calculation. Instead, the SMAP was allowed to compare all possible pockets in a pair of proteins that resulted in a global similarity score between the proteins. Higher similarities between members of a same target family (e.g. kinases, carbonic anhydrases etc.) show that the method can detect similarity in the protein binding pockets and provides a way of validation of the used method.

3.2.4 Networks Creation

In the drug-target network, the drugs and proteins are the nodes, which are connected by edges. The drugs and proteins nodes within the network are differentiated by their attributes or features associated with them. Similarly, the edges

connected between the nodes are also distinguished. The edges connecting nodes are distinguished into three types:

1. an edge between two drugs indicate that their fingerprints are similar.

2. an edge between two protein targets indicate that their 3D binding pockets are similar, and

3. an edge between a drug and a protein target indicates an experimentally known interaction that exists between a drug and a protein.

The Cytoscape [101] software environment (version 2.8.2) is used for the creation, visualization and analyses of drug-target networks. The drug-target interaction data is comma or tab-separated formatted for importing into Cytoscape. Cytoscape assigns unique identifiers to the interacting nodes and connects the nodes by edges that have interactions specified within the input file. Likewise, edges are also assigned unique identifiers. When the data for the network is imported, it resulted in a large drug-target network that comprises of 6,380 nodes and 57,064 edges. After importing the drug-target network, attributes for nodes and edges are imported from the tab-separated files exported from the HBP database subsequently. Drug and target nodes are distinguished by their attributes. Similarly, edge types are also distinguished by their attributes describing the interaction types (see above). Important attributes for nodes and edges are given in the following subsections.

3.2.4.1 Node Attributes

1. An attribute that classifies a node as a drug or target node if the attribute value is a string DRUG or GENE.

2. A disease attribute that classifies a drug for its indication area (e.g. attribute value is the disease name such as Alzheimer's disease or Epilepsy).

3. A status attribute that classifies a drug for its drug group, which distinguishes an approved drug from developmental or withdrawn drugs. The attribute value could be "approved", "discontinued", "Phase I" or so on.

4. A SMILE string is an important attribute for storing a drugs chemical structural data. The attribute value is a SMILE string. The SMILE attribute is used to visualize structures for selected drug nodes using CDK toolkit API.

5. An attribute classifying a target node in a certain pathway. The attribute value is the name of the pathway in which the gene is involved (e.g KEGG entity identifier).

6. A protein structure attribute for a target node is assigned if it has a PDB structure. The attribute values are the PDB ID(s) of the protein.

7. A mode of action (MOA) attribute is given for drug nodes if known.

3.2.4.2 Edge Attributes

1. An edge attribute that defines the interaction types into drug-target interaction, drug-drug similarity or target-target binding sites similarity.

2. An edge-score attribute is assigned as the interaction strength for drug-target interaction. The values are taken from the confidence score given in the STITCH database.

3. An edge-score attribute that shows the similarity value (Tanimoto Coefficient) between two drug structures.

4. An edge-score attribute that shows the SMAP similarity score between two target structures.

3.2.5 Pharmacome Tools Plugin

Cytoscape allows an easy-to-use API for developing plugins for performing specific analysis tasks over the networks. The "Pharmacome Tools" Cytoscape plugin was developed for the analysis and filtering of the NDD drug-target network. For our complex drug-target network, no specific plugin existed to perform the required specific tasks over the drug-target network. Therefore, a plugin was needed for the analyses of this complex network and the Pharmacome Tools plugin was created. This was developed in Java and can be used directly within Cytoscape for analyzing complex drug-target network containing different node and edge types. The CDK Toolkit [102] and MoSS (Molecular Substructure) Mining [103] are embedded in the plugin for providing molecular structure visualization and common substructure extraction. The following is a brief description of the major functionalities available within the Pharmacome Tools plugin:

1. Filtering polypharmacological drugs. Filtering is performed with the user specified thresholds values. Threshold values are considered for different edge types in edge score attribute. Filtering is initialized for drug pairs that

have similarity and interact with targets which also have similarities in the binding sites.

2. Filtering targets with polypharmacological profiles at a given target proteins pocket similarity. Threshold values are considered for different edge types in edge score attribute. Target filtering is initialized for target pairs that have binding site similarity along with their interacting drugs.

3. Prediction of new targets for existing drugs based on node neighborhood at the specified threshold values.

4. Extracting common structural scaffold for a set of selected drug nodes with the help of the MoSS API.

5. Visualization of chemical structures from the SMILE attributes using CDK toolkit API.

3.2.6 Toxicological Targets Prediction

To find whether a set of targets interacting with the failed or discontinued drugs is responsible for the drugs toxicity, a set of all interacting targets are retrieved for both sets of NDD drugs. A machine learning approach with recursive feature elimination (SVM-RFE) was applied to discriminate for the targets that interact with approved or discontinued drugs. An interaction matrix was generated that has the STITCH interaction scores between a drug and its target protein if there is an interaction present between the drug and its targets. Drug groups are put in rows and interacting targets are put in columns of this matrix. SVM-RFE will rank order those targets whose interaction is more common to one of the two drug groups.

3.3 Results and Discussion

The drug-target network visualizations and analyses are performed within the Cytoscape software environment. Figure 3.2 shows the complete and complex NDD drug-target network. Using Pharmacome Tools, the drug-target network is filtered and various specific views are generated which are analyzed for polypharmacological profiles of drugs or targets. Example of such views are clusters of specific drug groups or drugs targeting a specific disease.
In the following sections, different drug-target network analyses are performed to facilitate targeted polypharmacology. Through filtering a set of drugs and targets

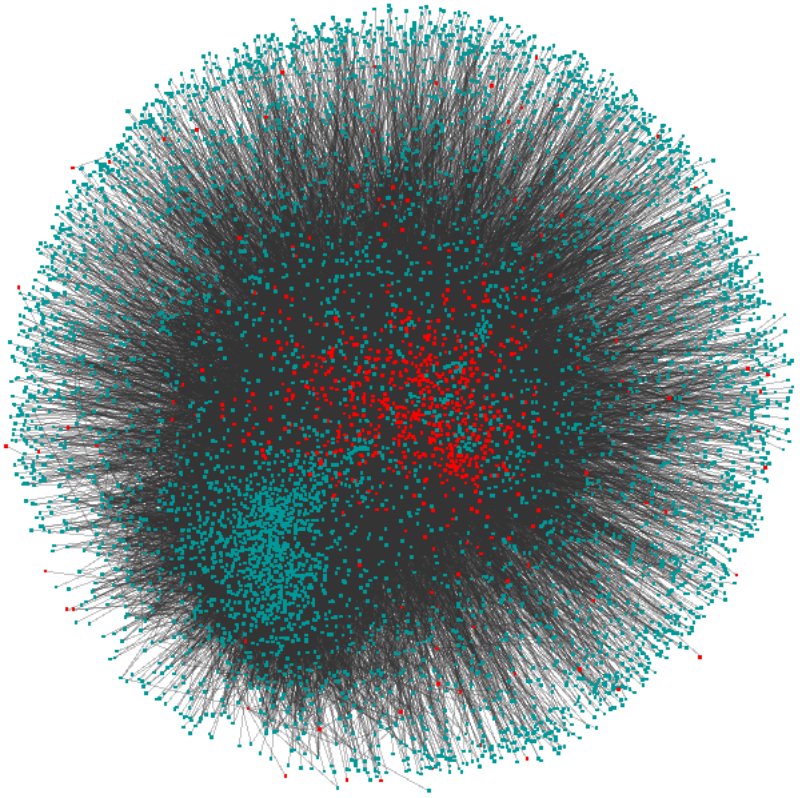


FIGURE 3.2: The drug-target network. Nodes in red color represent drugs and those in teal color represent nodes for target proteins. Black lines denote the edges connecting drugs or targets.

are selected for subsequent *in-silico* validation of polypharmacology by a virtual screening workflow that includes molecular docking and molecular dynamics simulations. The selected targets are also used for virtual screening against a focused compound library that is designed independently for a specific class of targets or a specific indication area.

3.3.1 Drug-Drug Similarity Network (Filtered By Indication Use)

The drug-drug similarity network shows that a large number of NDD drugs have similar chemical structures as they are connected by edges. Similarities also exist between drugs used for different NDD indications (e.g. drugs used for Alzheimer's disease have similarities to drugs used for Parkinson's disease or epilepsy). It is also not surprising that drugs used in one NDD indication area are undergoing clinical trials in another NDD indication area. Figure 3.3 shows that a large number of drugs that have therapeutic role in Epilepsy and Seizures have structural similarities with drugs that are used in Alzheimers disease and Parkinsons disease.

In this figure drugs used for different indication areas are distinguished by different colors and are connected by edges when they have structural similarities. This indicates that similar drugs used in different indications possibly interact with a common set of targets that are involved in multiple pathophysiological processes common to different NDD disorders.

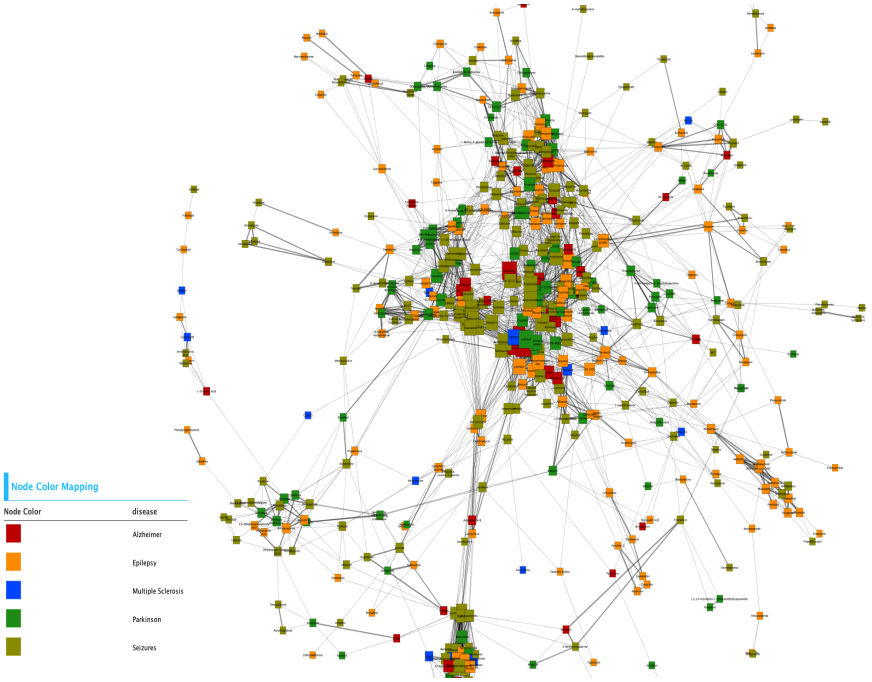


FIGURE 3.3: Drug-drug similarity network filtered by indication use. Color-coding indicates various NDD indications. Node size denotes the degree of the node. Edge thickness indicates the degree of chemical structural similarity.

3.3.2 Drug-Drug Similarity Network (Filtered By Drug Groups)

A network representation for drug groups is shown in Figure 3.4. In this figure, the FDA approved drugs are shown in red color, the developmental drugs in green color, and the drugs that are withdrawn or discontinued at various clinical phases are shown in yellow. It is interesting to note that the discontinued drugs have structural similarities to the approved or prescription drugs. If the discontinued and approved drugs have interactions with common target proteins, and also show structural similarities, then it will be more likely to predict side effects on the basis of structural similarity and interaction with similar targets. However, a limitation in this scenario is the lack of availability of side effects data for all drugs. Moreover, the interaction data is also not complete because the known experimental activities for both groups of drugs are incomplete. This network can be used for the study of

toxicity prediction or identification of toxic functional group in the developmental drugs if they have high similarities to the drugs that have adverse side effects.

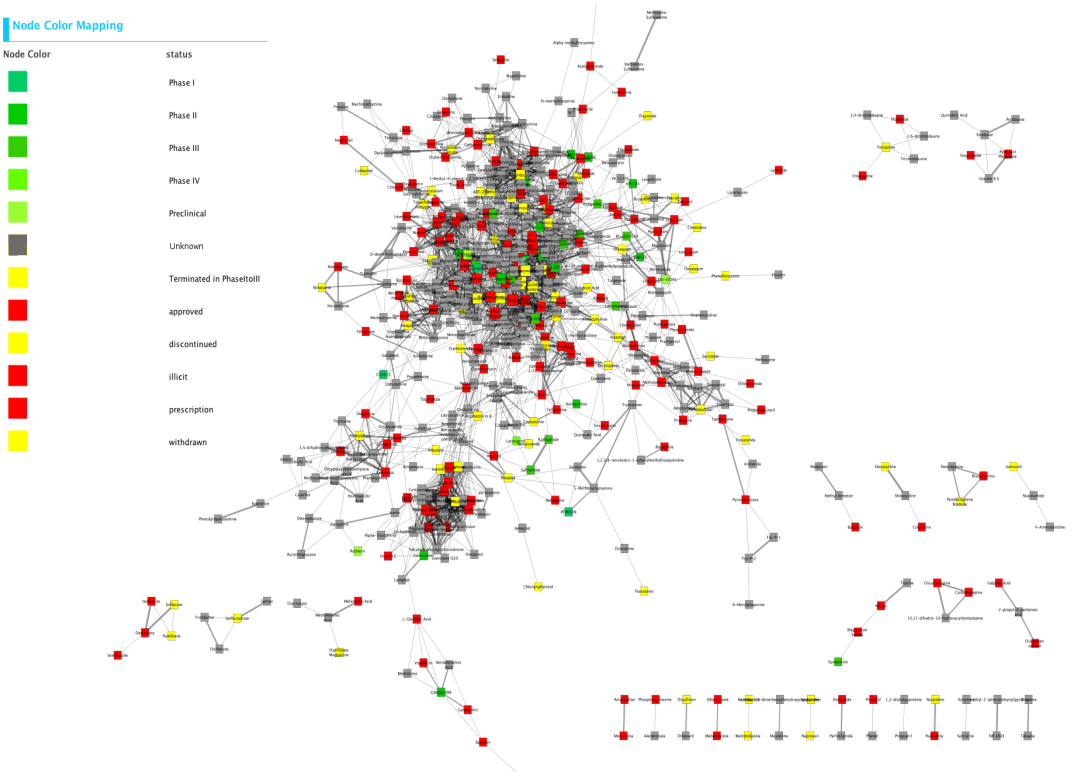


FIGURE 3.4: NDD drug-drug similarity network filtered by drug groups. Color-coding indicates drug groups. Edges indicate structural similarity between drug structures.

3.3.3 Network Clustering

Network clustering can be used to understand network architecture [104]. There are a number of algorithms that can perform network clustering on the basis of various attributes, but not all clustering algorithms can be applied that require edge weight attributes. Four of the clustering algorithms available in the Cytoscape plugin, clusterMaker [105] were used to perform clustering on the drug-target network. The complex NDD drug-target network is a bipartite network that contain different values for the same edge-score attribute. Values for drug-drug similarities are the Tanimoto Coefficients that are in the range between 0 and 1, while values for binding site similarities and confidence scores for interactions have variable ranges. Therefore, it is difficult to choose a cut-off value for the edge weight to perform clustering. Different results are achieved by applying different clustering algorithms. Figure 3.5 shows results of applying different clustering algorithms on the drug-target network. The Connected Components Clustering resulted in

a drug-target network that is very densely connected, since only one cluster produced. The "Community Clustering (GLay)" resulted in a total of 20 clusters, with three clusters being very large. The Affinity Propagation Clustering resulted in more than 200 clusters, which facilitate the isolation of drugs that target a single protein and proteins that are targeted by a single drug. The "MCODE Clustering" algorithm that finds highly interconnected regions in a network produced different size clusters that are neither very small as in the case of Affinity Propagation clustering, nor very large, as in the case of Community Clustering. The resulting clusters produced by MCODE Clustering isolate sub-networks, which may result in interesting patterns to be analyzed further. Therefore, this clustering algorithm should be used to isolate sub-networks that would facilitate the detailed study of small drug-target networks.

3.3.3.1 Drug-Drug Similarity Networks

Clustering was also performed on drug-drug similarity network that includes only drug nodes and their similarity edges. The network clustering can isolate clusters in which the drug nodes have high structural similarities as compared to drugs in other clusters. In order to observe the results of clusterings on this isolated drug-drug similarity network, the same clustering algorithms were applied as used in the drug-target network. Figure 3.6 shows the clustering results using the four clustering algorithms. The resulting sub-networks contain groups of very similar drugs that are used for different NDD indications. By specifying the chemical similarity threshold, variable number of clusters can be obtained that have high, medium or low structural similarities. One of the advantages of having different clusters of similar drugs is that one can isolate representative drugs from each cluster to have a diverse set of drugs that can be used as queries against a compound database in a similarity searching.

3.3.3.2 Target-Target Similarity Network

A sub-network of target proteins that have PDB structures available is isolated from the main drug-target network. The target proteins are connected when they have binding site similarities. The nodes within this sub-network are also differentiated by target types (i.e. therapeutic/successful targets, research and clinical trial targets as classified in the TTD [48] database). A large number of target proteins, which are either primary or secondary targets of NDD drugs, have similar binding pocket geometries. As shown in Figure 3.7, the similarity network is

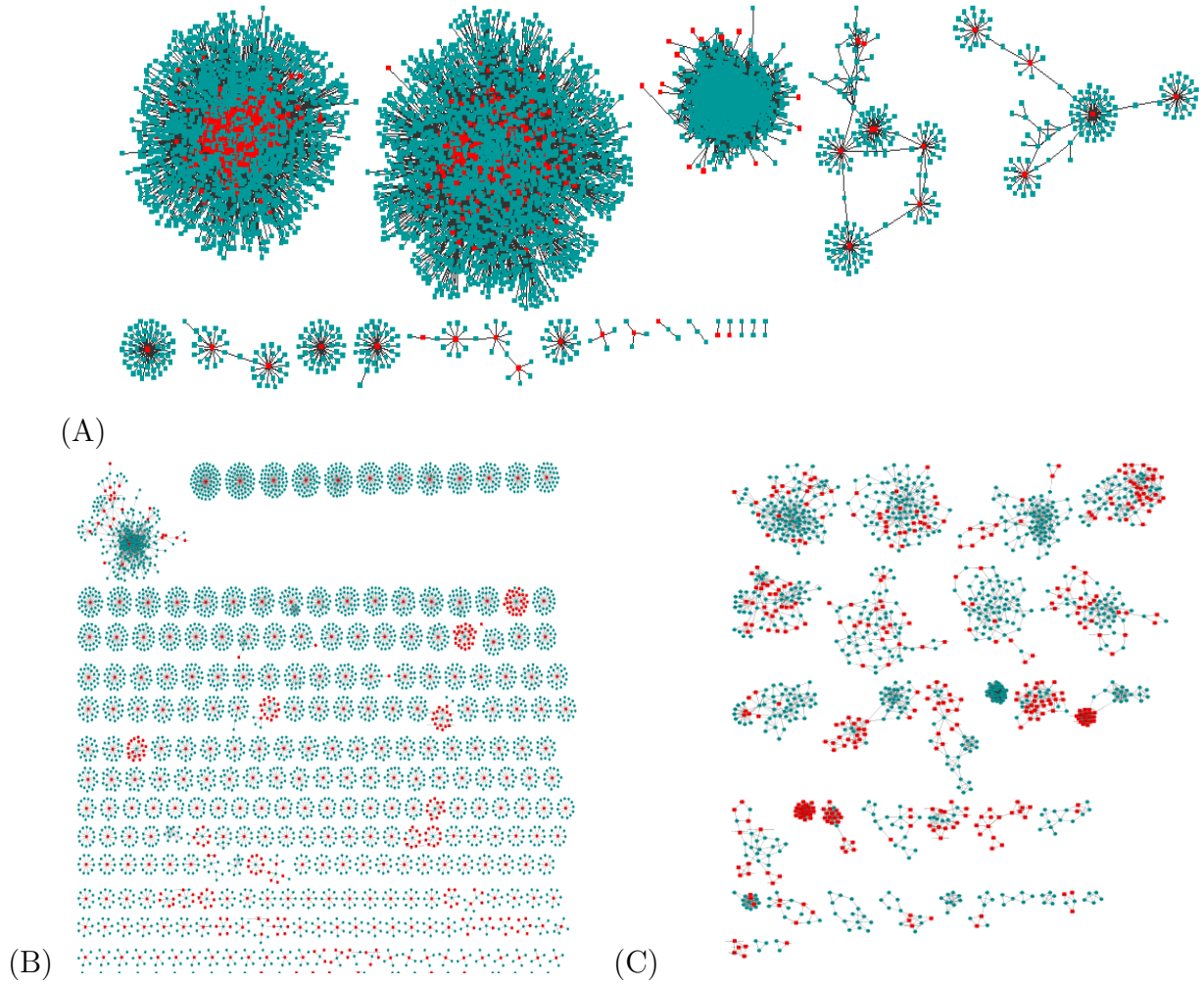
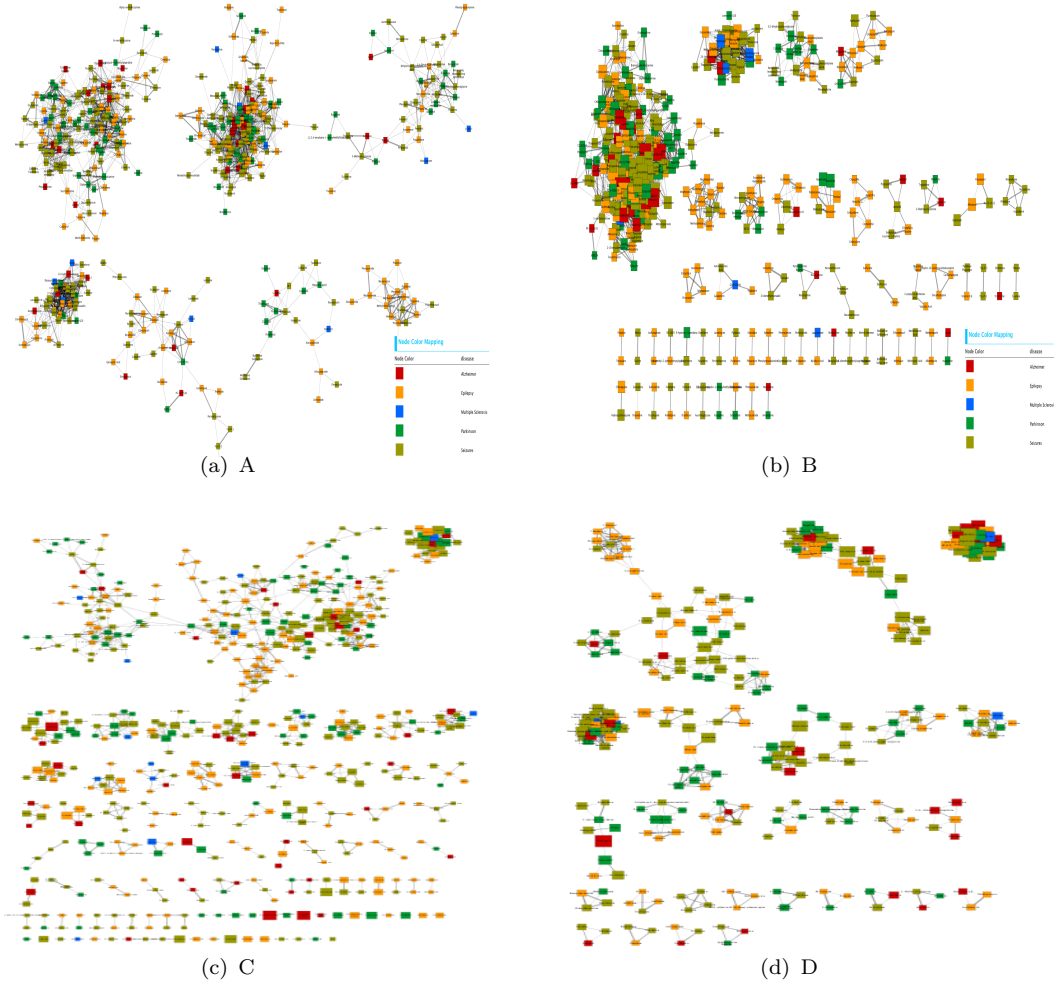


FIGURE 3.5: Results of three clustering algorithms applied on the drug-target network. (A) Community Cluster, (B: Affinity Propagation Clustering, (C): MCODE Clustering. Drugs are shown in red color and target proteins are shown in teal color.

densely interconnected that highlights some targets with high degree of connectivity. When the clustering algorithms mentioned above are applied to this network, they generate a few clusters that contain many targets having similarities. By varying level of similarity threshold (i.e. significant SMAP score) as encoded in the edge score attribute, clusters of targets with high or low similarities can be isolated. This binding site similarity sub network holds valuable information that can be utilized for designing highly selective or promiscuous drugs. A sub network may facilitate designing drugs rationally with desired polypharmacological profiles that target a group of proteins with similar pockets. Protein-protein physical interactions and knowing their function on the basis of similarity at the pathway level



(a) A (b) B

(c) C (d) D

FIGURE 3.6: Drug-drug similarity network clustering. Nodes are colored according to disease indications. (A): Community clustering, (B): Connected components clustering. (C): Affinity propagation clustering, (D): MCODE clustering.

will help elucidate the activation or de-activation of pathways that are involved in complex NDD disorders. Illustrating this idea in the subsequent sections, the protein similarity network is analyzed in combination with drug-drug similarity and drug-target interaction networks to identify patterns in the networks that can be used for the study of targeted polypharmacology.

3.3.4 Functionalities of Pharmacome Tools

3.3.4.1 Filtering Drug-Target Network

The drug-target interaction network is analyzed as a network pharmacology approach to study polypharmacology. The Pharmacome Tools plugin is used to filter drugs with polypharmacological profiles, filter the target proteins with similar

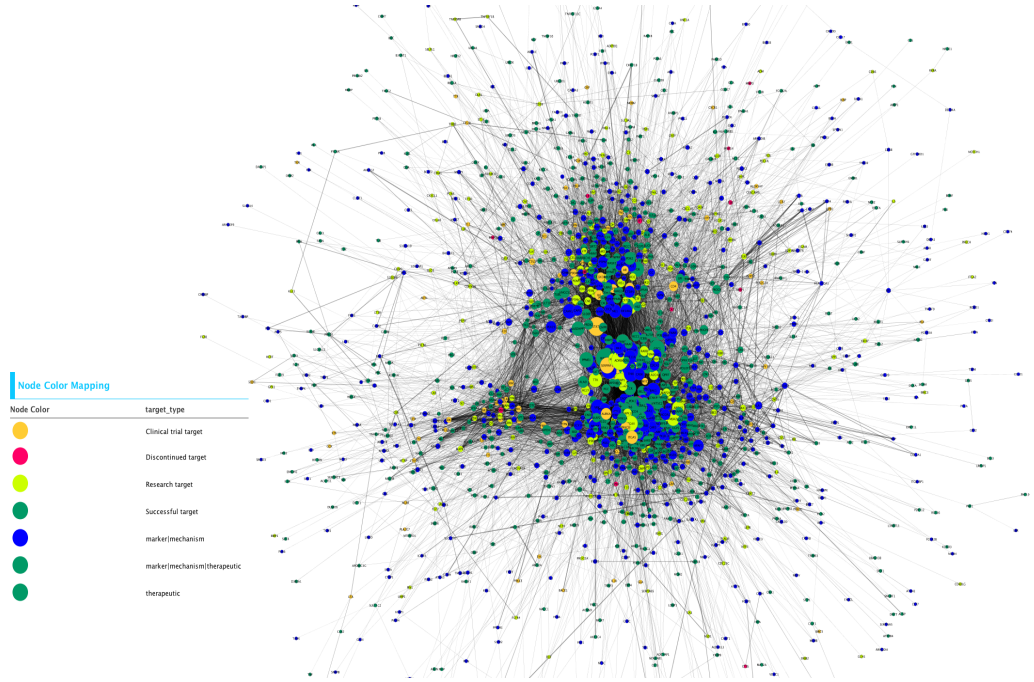


FIGURE 3.7: Protein-protein binding site similarity network. Node colors represent the target groups (e.g successful or research targets) and node size denotes the degree.

binding sites that can be targeted by promiscuous drugs, and predict new drug-target associations based on drug-drug chemical similarities as well as protein-protein binding site similarities. The algorithm used in the Pharmacome Tools plugin uses network neighborhoods identifying sets of drugs and targets with structural similarities in the presence of their respective drug-target interactions. The underlying idea is to first isolate drugs and targets that have high similarities that are extended by exploring the neighborhood of interacting drugs and targets. Likewise, polypharmacological filtering of target proteins is performed to isolate target nodes connected by similarity edges and their corresponding edge-related drug nodes. Subsequently, the resulting neighborhood of such sets of similar drugs or similar targets is explored for associations at the specified level of interaction strength. The STITCH confidence score for a drug-target association is taken as interaction strength. Ideally, protein-ligand binding activity data should be used as the interaction strength, but the lack of experimental data for most of the drug-target pairs impose this approximation. The following are some illustrative examples that demonstrate the use of Pharmacome Tools plugin.

3.3.4.2 Polypharmacological Filtering of Drugs

Studying the relationships between drugs and their targets at network level can serve as a useful indication on the polypharmacological behavior of the drugs. Figure 3.8 shows a view of a sub network that was filtered from the drug-target network with 6,083 nodes and 57,064 edges. The resulting filtered network contains only few nodes and edges (277 nodes and 925 edges). Three thresholds were specified that were used as the similarity cut-offs between drugs and proteins and confidence scores between a drug and a protein. Setting these three thresholds resulted in the filtered network as shown in Figure 3.8:

1. A similarity cutoff of TC ≥ 0.75 between drugs,
2. A SMAP similarity threshold ≥ 100 between target proteins and
3. An interaction strength cut-off of ≥ 0.90 between a drug and a target protein.

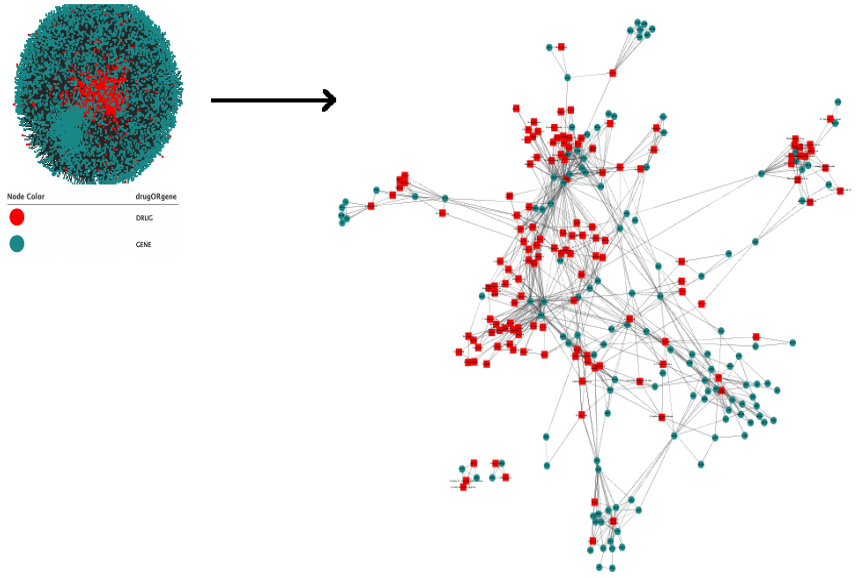


FIGURE 3.8: Polypharmacological filtering of drug-target network. (Left): The complete large network with 6083 nodes and 57,064 edges. (Right): The filtered network with 277 nodes and 925 edges. Filtering is done on the basis of three cut-off values that can be specified in the Pharmacome Tools.

Two approaches are used for combining the two-step filtering process. The first approach starts from isolating similar drugs and then explore the neighborhood of interacting target nodes. Those protein targets are isolated which are connected by similarity edges. The second approach starts from isolating similar protein targets and then explore the neighborhood of interacting drug nodes. The difference in the initial filtering for a set of similar drugs is that the drugs around the similar

target proteins need not be similar. However, a limitation in the second approach is the lack of structural data for the target proteins. Therefore, filtering is limited to the targets that have three-dimensional structural data available in the Protein Data Bank. The advantages in both filtering approaches are a) narrowing down the complex and large drug-target network to a smaller sub network that can be visualized and laid out more clearly than the original bigger network, b) allowing a more detailed study of targeted polypharmacology on a smaller set of highly similar drugs interacting with highly similar proteins and c) easing up selection of drugs or targets with the desired similarity values for virtual screening.

3.3.4.3 Drug Target Interaction Predictions

The systematic approach of predicting new drug-target associations based on drugs structural or proteins pocket similarities may allow to highlight drugs that have either polypharmacology or may lead to adverse effects. The Pharmacome Tools plugin can be used to infer new drug-target associations purely based on similarity. The new targets predicted for existing drugs would of course need experimental validation. Although, the drugs that bind to off-targets may introduce negative effects that could lead to toxicity and other undesired adverse effects [74], they should be studied for the registered side effects. If no adverse effects are reported for the drugs, it is more likely the drug will exert its effect through the desired polypharmacology. The newly predicted associations can be used to discover mode of action of successful drugs and propose new uses for existing drugs.
The new targets predicted by Pharmacome Tools are not based on the underlying network topology where associations can be predicted for the near neighbors of the nodes in the network. The plugin considers only the similarity between drugs and between the binding sites of their targets. The algorithm in the plugin first detects a set of drug nodes in the network that are connected by similarity edges with a user defined similarity threshold. For this set of interconnected similar drugs, all of the interacting target nodes are obtained that are above a given interaction cut-off. In the next step, pairs of target proteins are identified that are also interconnected by similarity edges with a given pocket similarity threshold. Based on the idea of "similar structure, similar property principle" [106], new edges are created between the pairs of drugs and target proteins if they are not already connected.
An example network is shown in Figure 3.9 (A) where two types of drug-target interactions are predicted. Similar to the polypharmacological filtering process described above, the prediction of new drug-target interactions is also based on two approaches. The first approach is the prediction of drug-target interactions

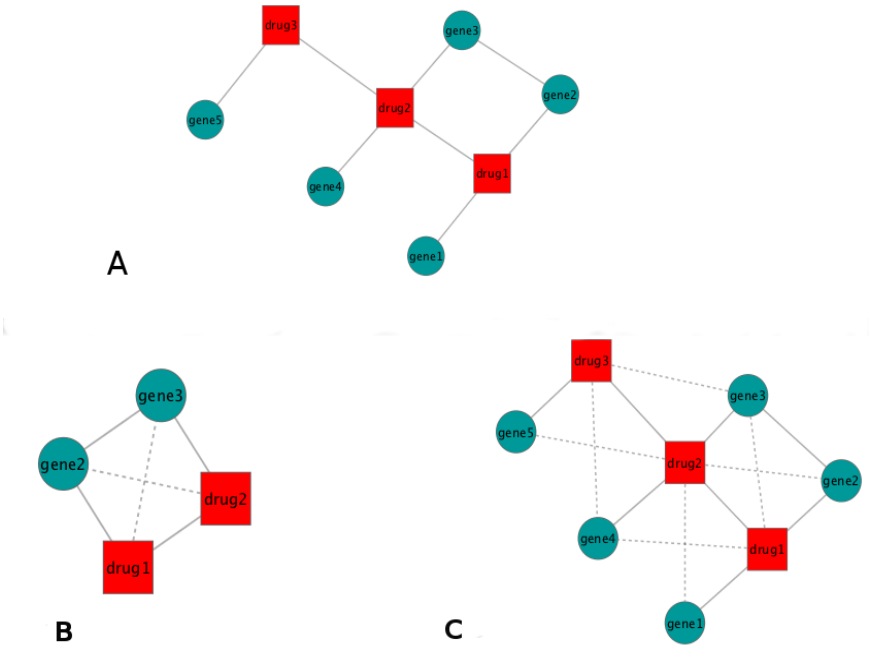


FIGURE 3.9: An example drug target network. Drug nodes are represented as rectangles in red color while target nodes are shown as green spheres. (A) Original network, (B) Predicted interactions for selected similar drugs interacting with similar target nodes only. The predicted interactions are shown as dotted lines. (C) Selected similar drugs with all interacting targets and predicted interactions are shown as dotted lines.

when both drug and target pairs are connected by similarities (Figure 3.9 (B)). The advantage in the first approach is that it isolates a sub network that includes the biologically meaningful predictions (since the interacting target nodes also have similar binding pockets). The new drug-target interactions that are predicted here are supported by evidences of similarities at both sides. However, a limitation here is the lack of three-dimensional structural data for the interacting proteins in the network, because not all protein nodes have structural data available and hence no pocket similarity results are available for them.

In the second approach, the new predictions are made when only the drug nodes are connected by edges and their interacting targets need not be connected by edges (Figure 3.9 (C). The advantage in this approach is that new associations can be predicted also for those proteins who do not have structural data due to which no similarities exist between those protein targets. However, a drawback in this approach will be producing large number of new drug-target associations that might include redundant or meaningless interactions (if we assume there exists no pocket similarity between the protein nodes) . But advantage over the first approach is that the target nodes are not limited to only those with available structures and more new interactions can be predicted. Redundant interactions

predictions can be avoided by increasing the thresholds for the drug-drug similarity and confidence score for the drug-target interactions.

3.3.4.4 Interactions Predictions for NDD Drugs

New drug-target interactions are predicted for NDD drugs present in the NDD drug-target network. There are total 6,083 nodes that include 724 drugs and 5,666 targets, which are connected by 57,064 drug-target interactions. The Tanimoto-Coefficient value ranges from 0.6 to 0.98 between the drug nodes, the pocket similarity scores ranges from 61 to 1050 between the protein nodes and the STITCH confidence scores ranges from 50 and 999 between drug and target nodes. Table 3.1 shows a number of new interactions predicted at different structural similarity thresholds among the drugs and the number of existing interactions already reported in STITCH database.

Table 3.1: New drug-target interactions predicted by Pharmacome Tools plugin for the NDD drug-target network. The predicted and existing interactions are given for various structural similarity thresholds.

Confidence = 0.80			No confidence		
Int. predicted	Int. in STITCH	TC cut-off	Int. predicted	Int. in STITCH	TC cut-off
4464	333	TC=0.60	17445	789	TC = 0.60
1284	136	TC=0.70	5090	349	TC = 0.70
888	118	TC=0.75	3221	290	TC = 0.75
559	89	TC=0.80	2048	210	TC = 0.80
289	45	TC=0.85	1091	114	TC = 0.85

The number of interactions that are already reported in the STITCH interaction database are between 10 to 15% when high confidence score for existing interaction is considered and a higher similarity threshold is specified. The r number of predictions can increase further if there are more structural data available for the targets where binding site similarities data can be added. In this approach, the Pharmacome Tools algorithm takes into consideration only those targets for which 3D structures are available and also have significant similarities in protein binding pockets. On the other hand, the number of reported interactions decreases (4 to 10%) when no confidence score for the existing interactions is considered. When only network topology is considered without taking protein binding pockets similarity into account, the number of possibilities to predict new interactions will increase further, because the target set is not limited to only those with available

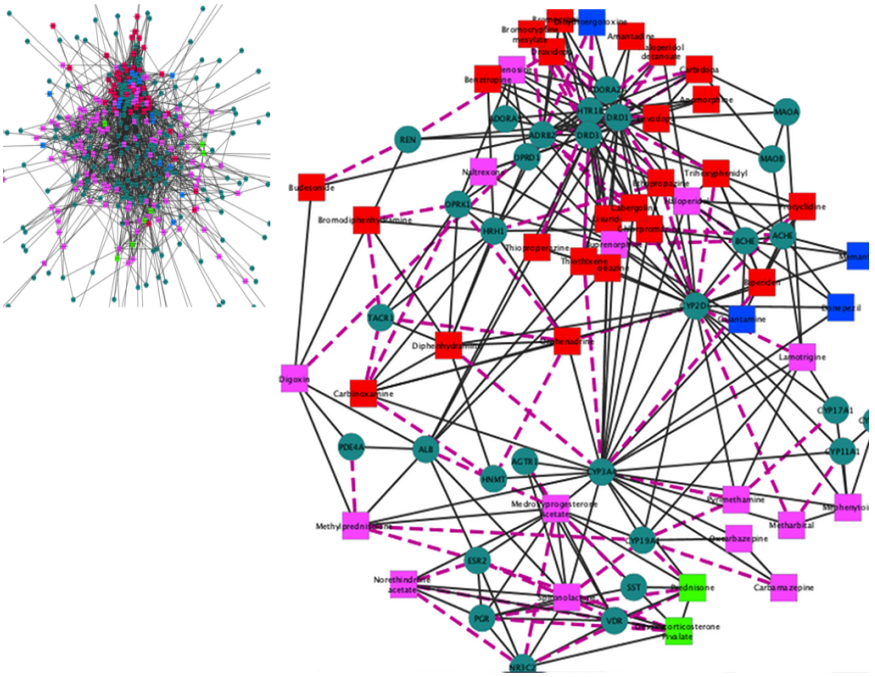


FIGURE 3.10: Interaction predictions in a sub-network of approved drugs interacting with approved therapeutic targets only. Target nodes are colored in teal. Drug nodes are colored according to disease indications, Red: Parkinson's disease, Blue: Alzheimers disease, Pink: Epilepsy, Green: Multiple Sclerosis. Predicted interactions are drawn with a dotted line in pink.

structures. Interesting interactions could be predicted by limiting the drug nodes to approved drugs that are only interacting with established targets of therapeutic relevance.

Figure 3.10 shows some of the new interactions predicted for approved drugs that are used in multiple NDD indications. These interactions are predicted on the basis of similarity between both drugs and targets. The similarity threshold for drugs was set to TC ≥ 0.7 and for the targets the SMAP score above 100. As a way to provide improved predictions based on more rigorous *in-silico* methods such as cross-docking and molecular dynamics simulations, a few of these interactions have been studied, which will be described in detail in a separate chapter. Such predicted new interactions can be tested experimentally for validation.

3.3.4.5 Polypharmacological Scaffold Detection

Molecular Substructure (MoSS) algorithm is integrated in the Pharmacome Tools plugin that can help extracting a common molecular substructure present in a set of drugs acting against multiple targets. The common chemical scaffold can be visualized over the network with the help of CDK Toolkit. The obvious requirement is that molecular structures for the drugs are provided.

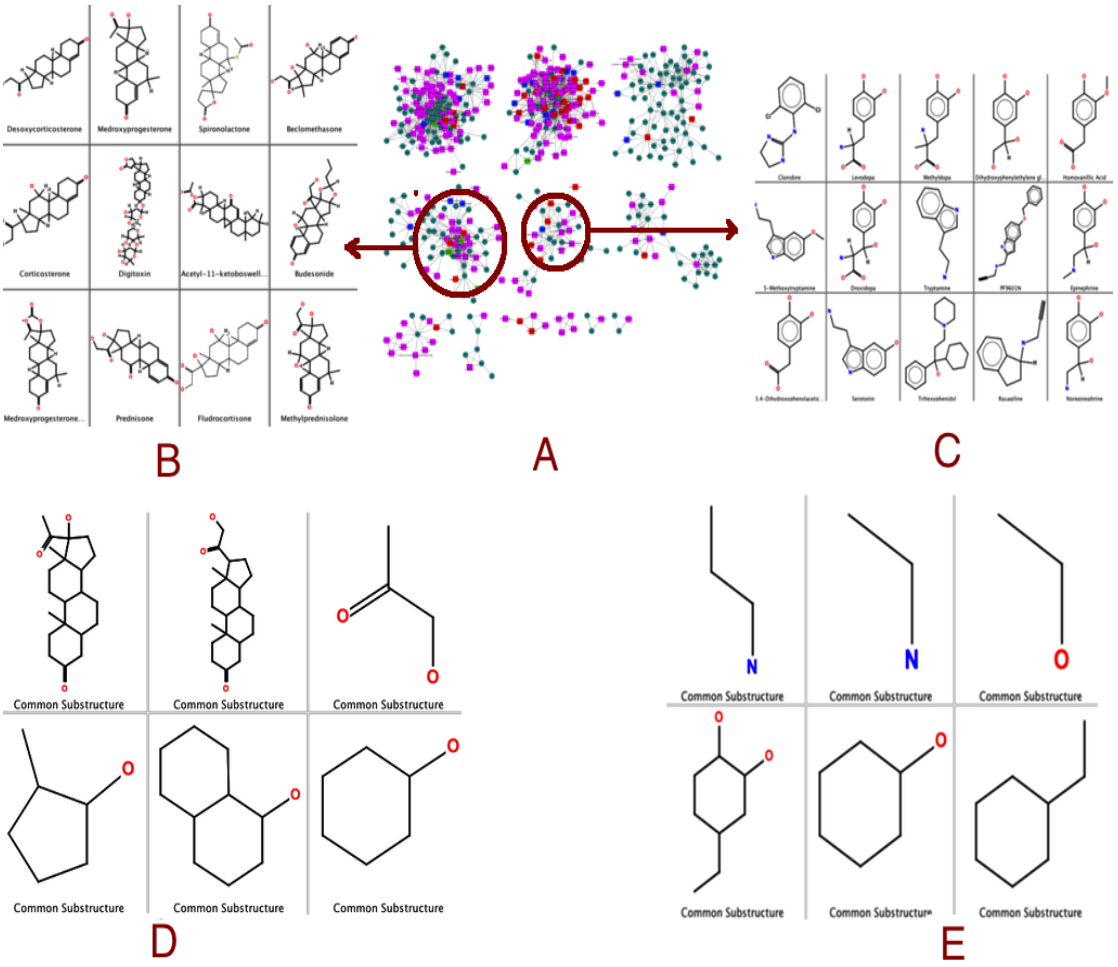


FIGURE 3.11: Polypharmacological scaffold extraction. A: Clusters of drug-target network as visulaized by community clustering algorithm. B: 2D structural representations of multiple drugs in one cluster. C: 2D structural representation of multipe drugs in another cluster. D: Common molecular substructures detected for drugs shown in B. E: Common molecular substructures detected for drugs shown in C.

The common molecular scaffold present in the drug structures may reveal the polypharmacological behavior of those drugs. Exploration of such promiscuous chemotypes could help guide the design and evaluation of multi-target drugs. A promiscuous substructure present in a single drug that is active against multiple targets that do not have similarities at the binding site could also show its polypharmacological behavior, although the drug might bind/interact differently in a different protein binding site by adapting different conformation. To visualize common skeletal scaffold in a set of drugs interacting with multiple targets, clustering can be applied over the drug-target network to isolate groups of drugs and multiple targets. Figure 3.11 shows two examples of polypharmacological scaffold extraction for two sets of drugs selected from two clusters. These drugs are

clustered and grouped together with their multiple interacting targets by the Community Clustering algorithm. Molecular substructures that are frequently present in a set of drugs structures are extracted and visualized (Figure 3.11 D and E).

3.3.4.6 Selection of Targets for Polypharmacology Study

With the help of Pharmacome Tools, important protein targets are selected with a focus on targeted polypharmacology. As the drug-target network is very large, including thousands of protein targets and hundreds of established therapeutic targets, it is a key issue to select a few relevant targets for virtual screening. Appropriate target selection is one of the first steps in virtual screening prior to molecular docking and molecular dynamics. To facilitate the study of targeted polypharmacology, a set of targets is selected on the basis of connectivity in the drug-target network. The connectivity is presented by binding site similarities and interaction with approved drugs around these targets that are used in various NDD indications. Target selection is facilitated by the polypharmacological filtering algorithm which is one of the useful features of Pharmacome Tools as described in the previous section.

The rational approach towards polypharmacological target selection includes consideration of binding site similarities and interaction with approved drugs used in multiple NDD indications. Therefore, the drug-target network is pre-filtered to include only established targets and those targeted by approved drugs. Figure 3.12 (A) shows some of the candidate targets that are considered useful for the study of targeted polypharmacology based on their binding sites similarities with each other as well interaction with drugs that are used for Alzheimers disease, Parkinsons disease and epilepsy. Furthermore, these targets also make a visible cluster in the target-target similarity network, which highlights a higher similarity between them and similarities to many other target proteins (Figure 3.12 (B)). Those proteins, which have only X-ray structures are considered. Following is a list of selected targets for virtual screening:

1. beta-2 Adrenergic receptor (ADRB2)
2. Dopamine D3 receptor (DRD3)
3. Adenosine A2A receptor (ADORA2A)
4. Histamine H1 receptor (HRH1)
5. Chemokine C4 receptor (CXCR4)

6. Mono-amine Oxidase B (MAOB)

7. Phosphodiesterase 4A (PDE4A)

3.3.5 Toxic Targets Prediciton for Discontinued Drugs

The set of discontinued NDD drugs share a set of common targets that are almost not interacting with the set of approved drugs used for NDDs. These results shown in Figure 3.13 demonstrate that most of the discontinued drugs share interactions with a set of targets. While these interactions are not seen with the set of approved drugs used in NDD, these observations for the discontinued drugs provide a possible reason for their toxicity that ultimately lead to their discontinuation.

3.4 Conclusions

A demonstration of the use of HBP for visualization and analysis through *in-silico* network pharmacology approaches is presented in this chapter. Using network approaches enable us to generate a system level view of the specific knowledge that exists in the HBP knowledge base. Extending the knowledge base to incorporate additional data further increases the complexity of the underlying networks. It was observed that even an NDD pharmacome focused part of the HBP resulted in a large, complex network of drug-target interactions. Such a complex NDD drug-target network is analyzed from various aspects. Using network algorithms can reduce the network complexity of drug-target networks. The "Pharmacome Tools" plugin is developed to reduce the complex NDD drug-target network by filtering a set of targets and drugs that are potential candidates for the study of polypharmacology. The plugin is also able to predict new drug-target associations that could be investigated for the prediction of off-target effects and the repurposing of existing drugs. Functionalities of the Pharmacome Tools plugin were used to filter the large network, obtain a relevant data set for workflow validation and to select a highly prioritized set of protein targets for the study of targeted polypharmacology. A weakness of this approach to select protein targets is that enough structural data for the protein targets should be available. However, using the available structural data, a strong point of the approach is that one can still be able to identify and rationally select a set of protein targets that can be used in the design of multi-target drugs.

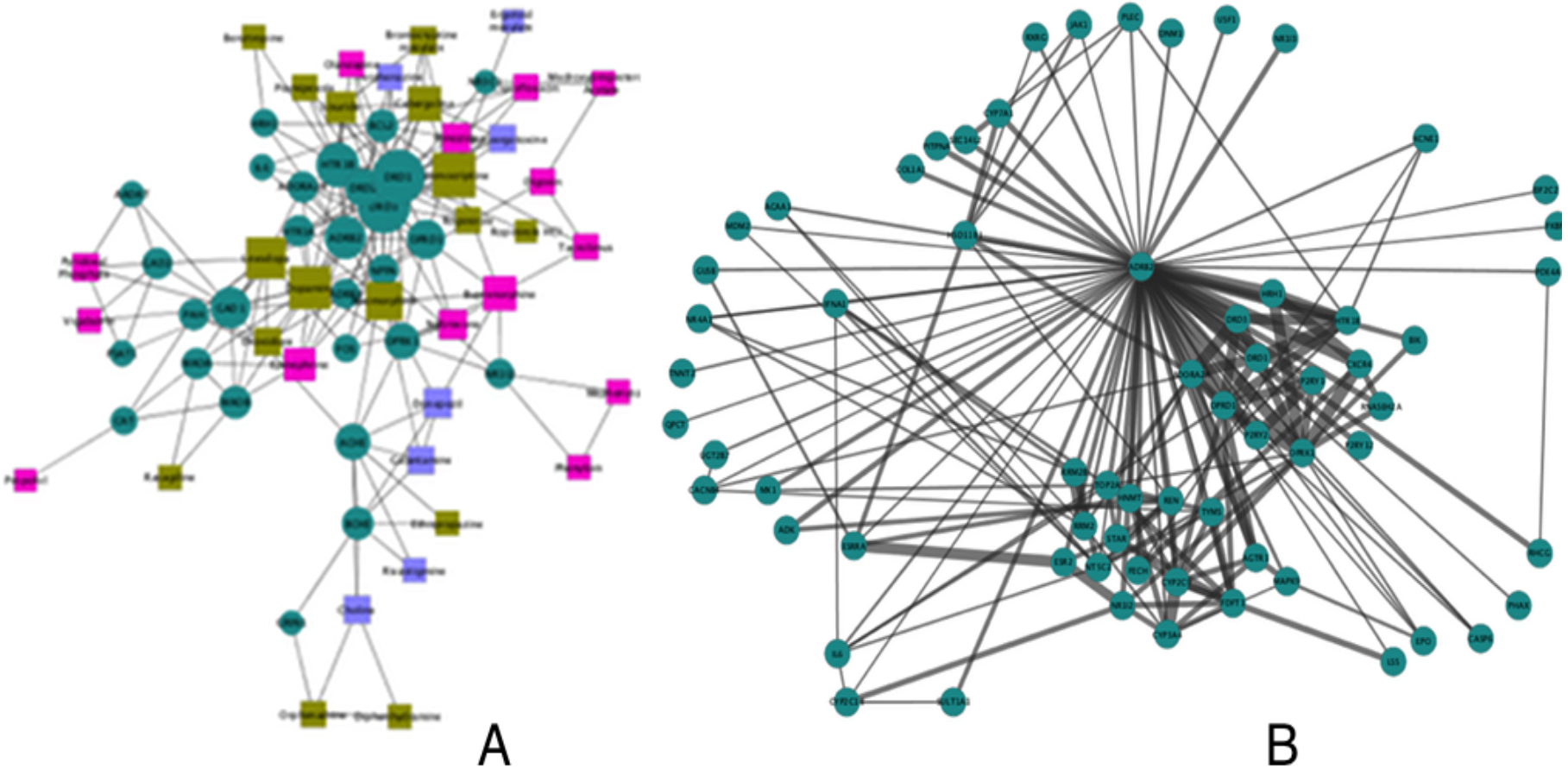


FIGURE 3.12: Polypharmacology based selection of targets. A: a sub-network of multiple targets with approved drugs for multiple NDD indications. Drug nodes are colored by disease indications; anti-epileptic drugs are shown in pink, anti-alzheimer drugs in light blue and anti-parkinson drugs in brown. B: Protein pocket similarity network highlighting the targets with higher similarities at binding pockets. Edge-thickness denotes higher binding site similarity between the protein nodes.

FIGURE 3.13: Heatmap of drug-target interactions prioritized for discontinued drugs. Drugs sets (approved; above, discontinued; below) are given in rows and their interacting targets are given in columns. Horizontal dashed line divides the drug groups into known FDA approved known drugs and discontinued drugs, the arbitrary vertical dashed line highlights the interacting targets common to the set of discontinued drugs.

Chapter 4

SVM based Classification of Neurodegenerative Disease Drugs for Pharmacological Modeling

4.1 Introduction

The enormous complexity of the brain and the poor understanding of neurodegenerative disease mechanism challenges the success rate of drug discovery efforts. These challenges include pharmacokinetic hurdles of therapeutics to reach the brain and their side effects, impacting drugs development for neurodegenerative disorders (NDD) [107]. Another potential complication is the involvement of multiple targets in a single disease related pathway. Several efforts have been undertaken to develop targeting strategies that take into the account the complex pharmacology of brain disorders, including dual the targeting of MAOB and A2A receptors in Parkinsons disease [108], multi-targeting of amyloidogenesis[109] and dual inhibition of AChE and serotonin transporter in Alzheimers disease [110]. However, the rational design of ligands acting on multiple targets requires that such ligands bear certain molecular properties that are represented by molecular descriptors.

As the descriptor (i.e. feature) space representing these drugs is very large, the calculation and analysis of all possible descriptors for identifying similar drugs is computationally expensive. Therefore, machine learning based statistical methods are well suited to reduce the dimensionality and thus limit the search space. Learning methods have been widely used in computational chemistry to facilitate predictions of pharmacokinetic and toxicological properties of chemical agents by

employing a variety of molecular descriptors to characterize structural and physiochemical properties of drug molecules [111]. Proper selection of relevant molecular descriptors using machine-learning methods could be useful for predicting drugs that treat central nervous system (CNS) disorders [112].
However, selecting an optimal feature subset is an NP-hard optimization problem [113]. Existing feature selection methods can be categorized into two classes: filter methods and wrapper methods. Filter methods, such as Principal Component Analysis (PCA), rank the features in a rather unsupervised way by using heuristics based on general data characteristics. Wrapper methods, including a variety of Support Vector Machines (SVMs), use induction algorithms itself and iteratively select features by evaluating their usefulness for the target predictor. Both methods have been already applied to identify drug-like molecules from non-drug molecules using molecular descriptors [114]. In a recent study, Tang et al. [115] compared four machine-learning methods to discriminate approved drugs from experimental ones and showed that SVM models performed best in their test scenarios. Use of Recursive Feature Elimination (RFE) combined with statistical learning methods to select the proper set of molecular descriptors has shown to substantially improve the discrimination of brain drugs from non-brain drugs [6]. Similarly, SVM based classification of CNS drugs based on their molecular descriptors has been reported to outperform other statistical learning methods in selected scenarios [116]. Our study focuses on the application of this well-established machine learning approach for characterizing approved drugs in the NDD area with the goal of being able to differentiate NDD-specific drugs from non-NDD ones.
In order to analyse and distinguish molecular properties that discriminate NDD-specific drugs from non-NDD ones, we inspected a set of approved drugs for NDDs in search of specific molecular descriptors. These molecular descriptors may contribute to the established safety and efficacy profiles of these drugs as they are approved for use in CNS. For this purpose, we applied the SVM-RFE method [117]. SVM-RFE is a widely adopted state-of-the-art supervised feature subset selection approach. It has been previously applied with great success to gene selection tasks in microarray expression data [118]. Using this method, we ranked the features according to their contributions to define the maximally separating hyper plane. Our analyses show high classification accuracy in classifying NDD drugs from non-NDD drugs. Higher classification accuracy is achieved with SVM-RFE as compared to conventional similarity measures and a simple supervised feature selection and classification algorithm. We propose to use the predictions made by this classifier to design focused libraries for the discovery of novel NDD drugs.

4.2 Methods and Material

The drug data set was compiled carefully before calculating molecular descriptors. The data set comprised of 160 FDA approved drugs used in NDD disorders such as Parkinsons disease, Alzheimers diseases, epilepsy and multiple sclerosis, were collected from various knowledge base resources including DrugBank [**?**], CTD [49], TTD [48] and STITCH DB [60]. An additional set of 253 FDA approved non-NDD drugs were randomly collected from DrugBank and included in the data set. A set of 314 molecular descriptors was calculated using MOEs QuaSAR-Descriptor function [92]. The drug molecules were prepared before calculating descriptors. Preparation of drug structures included the generation of 3D conformations, removing salts and counter-ions and getting optimized 3D geometries via applying energy minimization with MMFF94x force field [119]. The final descriptor space included all 2D descriptors that use only atomic and connectivity properties of the molecules and all i3D descriptors that use internal spatial information about each molecule (e.g. potential energy, volume and solvent accessible surface area).
A pairwise correlation matrix was generated for the descriptors to identify highly correlated set of descriptors. The data set was grouped into two classes and labelled as NDD+ for NDD drugs and NDD- for non-NDD drugs. The data set was then partitioned by random sampling into a training set (70%) for machine learning model selection and a test set (30%) for model assessment.
Molecular descriptors are ranked and selected based on the SVM machine learning method with recursive feature elimination (SVM-RFE), which utilizes the sensitivity analysis based on the weights of support vectors in each descriptor dimension. Given a training set $\{(\mathrm{x}_i, \mathrm{y}_i) \quad — \quad \mathrm{x}_i \quad \epsilon\, \mathrm{R}_p, \mathrm{y}_i \quad \epsilon \quad \{+1, 1\} \quad \mathrm{i} = 1, 2, \ \mathrm{N}\}$ of N instances, an SVM determines an optimal decision function of the form:

$$f(x) = \sum a_i.y_i.K(x, x_i) \tag{4.1}$$

Where a_i can be interpreted as a weight of each x_i and K is any kernel function. SVM-RFE iteratively eliminates features by using the equation of maximally separating hyper-plane:

$$w = \sum a_i.y_i.x_i \tag{4.2}$$

and removing features corresponding to the dimension *argmin* (w_i^2), j=1,2,.... p. Subsequently, based on those selected descriptors, an SVM (R implementation in the e1071 package) is trained to classify NDD drugs from non-NDD drugs. We observe optimal classification performance when using the Radial Basis Function

(RBF) kernel. In order to stabilize and establish the robustness of the resulting feature rankings, the resampling technique with cross validation, Multiple SVM-RFE is used [120]. A generalization error (by 10 fold cross validations) is also estimated for a number of top ranking descriptors, which shows the performance of classification based on the number of selected ranked features.

As a comparison, three molecular similarity based approaches were evaluated to compare the performance of the classification based on similarity measures. 1) The ShaEP algorithm [121], which calculates molecular electrostatic potential based descriptors. 2) Feature trees (FTrees) algorithm [122], which represent chemical properties of molecules by a set of features. 3) MOEs similarity search using MACCS [123] fingerprints. ShaEP performs a rigid-body superimposition of 3D molecular models that is achieved by finding similarities in the molecular electrostatic potential (MEP) field followed by volume overlap. The drug structures were energy minimized with MMFF94x force field and their 3D coordinates were saved as Tripos Mol2 file format. It is necessary to use Mol2 format because it supports partial atomic charges calculated for all atoms in a molecule that are needed for computing the molecular electrostatic potential. Data sets composed of drugs active against specific target families (e.g. beta-2 adrenergic receptor (ADRB2), estrogen-related receptor alpha (ESRRA) or acetylcholinesterase (AChE) interacting drugs) and combined clusters were used as queries against all approved drugs to recover the same drugs by using ShaEP similarities comparisons. Similarities calculations were performed for all individual NDD approved drugs against the non-NDD drugs. Average similarities across all NDD drugs were considered for enrichment analysis and calculation of ROC curves.

Additionally, a simple supervised feature selection and classification algorithm based on linear discriminant analysis (LDA) [124] was also used as a comparison to SVM-RFE. LDA seeks to reduce dimensionality while preserving class discriminatory information in terms of features. It is known that the t-score is a natural and optimal feature selection criteria in LDA if no correlation among features is observed [125]. This special case of LDA is known as Diagonal Discriminant Analysis (DDA). In the fields of gene selection and ranking, different LDA based approaches have been successfully used for gene ranking and selection [126], [127]. These approaches usually consider correlation among features. We used state-of-the-art LDA feature selection approach that is based on correlation adjusted t-scores (CAT scores) [127]. Given that the training data belonging to two classes of sizes n1 and n2 and with means μ_1 and μ_2, feature specific CAT scores vector τ^{adj} is defined as:

$$\tau^{adj} = P^{\frac{-1}{2}}\{(\frac{1}{n_1} + \frac{1}{n_2})V\}^{\frac{-1}{2}}(\mu_1 - \mu_2) \tag{4.3}$$

where P = (ρ_{ij}) is the correlation matrix and V = diag$\{\sigma 12$, $\sigma 22$, . . . , σp2$\}$ is a diagonal matrix containing the variances. Both P and V are obtained from decomposition of the common covariance matrix $\sum$ such that

$$\sum = \mathrm{V}^{\frac{1}{2}}\ \mathrm{PV}^{\frac{1}{2}}.$$

Notice that when there is no correlation among features, that is when P = I, τ^{adj} reduces to a multiple of $\mathrm{V}^{\frac{-1}{2}}(\mu_1 \quad \mu_2)$ which represents two sample t-score. Thus definition in equation 4.3 allows a mechanism for feature ranking based on DDA and LDA.

4.3 Results

We considered Pearson correlations as a measure to identify non-correlated set of descriptors. The pairwise correlation matrix of descriptors (a subset is shown in Figure 4.1) shows that a large number of descriptors are inter-correlated. Therefore, only one of the descriptors in the correlated clusters of descriptors was randomly selected and the rest were discarded if highly correlated (i.e. Pearson correlation coefficient ± 0.8). Through this elimination procedure, a set of 107 non-correlated descriptors was identified for subsequent analysis.

4.3.1 SVM based Descriptor Selection and Classification

A large number of molecular descriptors represent the drugs in the multi-dimensional feature space. Not all of these descriptors are equally discriminative in classifying NDD and non-NDD drugs. We notice, in an unsupervised fashion, that many of these descriptors are highly (in case of Pearson correlation) correlated and therefore, will not bring additional information if we use them together. Since the descriptors may bear even complex relationship among each other, we employ a supervised descriptor selection and classification method to determine relevant descriptors with respect to their importance towards data classification. To this end, SVM-RFE is one of the state-of-the-art methods that recursively eliminates the less relevant features and selects the most relevant features best describing the data. Table 4.2 shows a list of top 10 ranking descriptors for our NDD drug data set, ranked by SVM-RFE, on the basis of which we are able to discriminate and classify NDD drugs from non-NDD drugs. The set of 40 top ranking descriptors

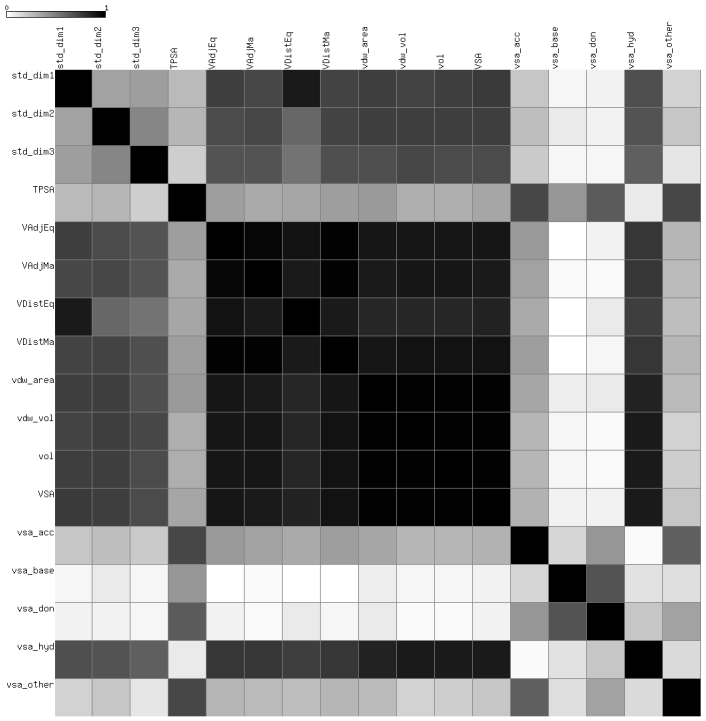


FIGURE 4.1: A sub-matrix of the correlation coefficients matrix showing highly positive or highly negative (in dark) correlations among a set of molecular descriptors. Molecular descriptors are given in rows and columns. Absolute values are taken for the negative correlation values.

(based on lowest generalization error estimate as shown in Figure 4.2) is used to train an SVM classifier for our data set. The training data set included only NDD drugs combined with drugs for disorders other than NDD. To see whether a set of descriptors would be relevant, the classification was performed using three sets of descriptors: all descriptors (total 314), non-correlated descriptors (total 107) and top ranked selected descriptors (top 40). The prediction performance was evaluated with ten fold cross validation of the underlying classification models for the three sets of descriptors. Classification results have overall accuracy of about 80% for the top ranked selected set of descriptors.

TABLE 4.1: Classification accuracies (10x CV) of SMV-RFE and LDA based approaches for NDD data set using: 1.all descriptors (314), 2. NC (non-correlated 107) and 3. TR (Top Ranked 40, by SVM-RFE and LDA.

Descriptors sets	SVM Accuracy %		LDA Accuracy %	
	Training	Test	Training	Test
All	82.7	83.9	75.6	74.8
NC	83.0	81.5	74.4	71.5
TR	82.4	79.8	70.3	72.4

Table 4.1 shows the results of classification accuracy using three sets of descriptors for our data set comprising NDD drugs mixed with other drugs for other disorders. The results obtained with SVM are better than the LDA based method. The classification performance obtained with using non-correlated and a few selected descriptors is nearly the same as that of considering all descriptors. Therefore, if a limited number of descriptors can produce comparable results, the advantage will be efficiently computing a small number of relevant descriptors for a large data set of molecules in a compound library.

TABLE 4.2: List of top 10 ranking descriptors that contribute to classification and predictions of NDD drugs

No	Descriptor code	Description	Avg.Rank[1]
1	E_sol	Solvation energy	1.36
2	chiral_u	Unconstrained chiral center	4.18
3	b_double	No. of double bonds	13.14
4	PEOE_VSA-4	Partial charge/VDW Surface Area (bin-4)	16.54
5	SlogP_VSA8	Log of octanol/water partition coefficient with subdivided surface area, (bin-8)	18.72
6	DCASA	Difference between CASA+ (charge weighted +ve surface area) and CASA-	19.18
7	vsurf_IW7	Hydrophilic INTEGY (INTEraction enerGY) moment (bin-7)	21.88
8	SlogP_VSA2	Log of octanol/water partition coefficient with subdivided surface area, (bin-2)	22.76
9	SMR_VSA4	Molar refractivity with subdivided surface area (bin-4)	22.78
10	vsurf_HL2	Hydrophilic-Lipophilic(bin-2)	23.20

[1] Average rank across 50 iterations for descriptors ranking and selection

Table 4.2 enlists the top ranked descriptors, which were used for classification. These ranked descriptors include the most widely used descriptors in QSAR studies (e.g. logP and surface area based descriptors) which describes the molecules lipophilic or hydrophobic properties and are valuable parameters for determining the ability of drugs that should be absorbed and penetrated into the brain.
These results show that there are sets of molecular descriptors that can be used to discriminate NDD drugs from non-NDD drugs. The underlying statistical learning methods also provide a way of limiting the number of ranked features based on generalization error estimation. This is helpful because calculating these molecular

descriptors is compute-intensive and would require a substantial amount of computation time to calculate descriptors for a large number of molecule structures in a compound database. It is, therefore, desirable to select a minimum number of top ranked descriptors with lowest generalization error. Figure 4.2 illustrates a generalization error estimate for a number of ranked descriptors. An independent test data set that was not included in the feature selection was used for calculating the generalization error. The generalization error estimate was calculated for top 50 descriptors only because calculating it for all features will increase the number of underlying RBF kernel SVMs and hence would require a substantial amount of computational time. The results indicate that, when compared to using all 314 molecular descriptors, a similar accuracy can be achieved by using only a few relevant and highly discriminative descriptors selected by SVM-RFE and with decreased number of support vectors for the SVM model.

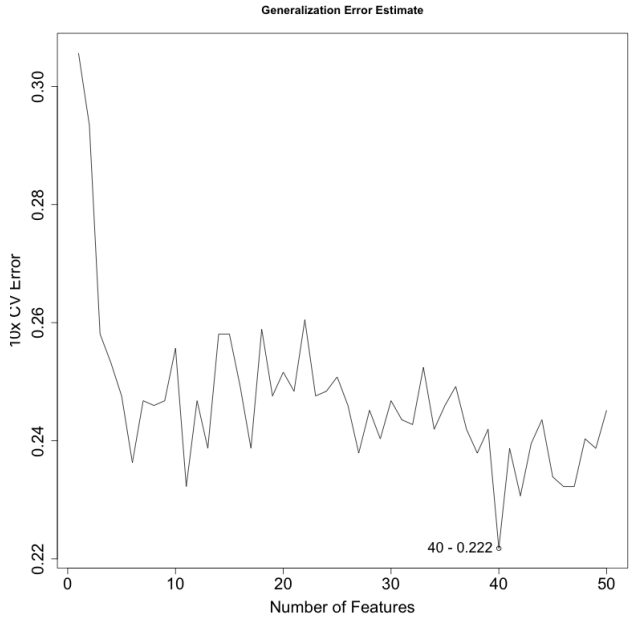


FIGURE 4.2: Plot of generalization error against a number of ranked descriptors over the independent test set. The 10x cross validation error estimate shows that the top 40 descriptors result in higher classification performance.

4.3.2 Similarity Searching

The similarity results obtained by three similarity measures are given Figure 4.3. Mean ROC-AUC (computed with the help of an R package caTools) of 0.60 was obtained for similarity comparison using ShaEP, 0.56 for similarity results of FTrees and 0.52 for similarity using fingerprints. These small ROC-AUC values are close

to random suggesting that a classification strategy based on only structural similarities should not be used in classifying a set of NDD drugs from non-NDD drugs. Running target family specific clusters of drugs as a query against the remaining NDD drug data set showed that we could not get a reasonable similarity among the individual clusters of drugs. The reason for this could be that there are drugs with common structural properties targeting a specific protein family and different drugs have variable binding affinities towards different protein targets. Therefore, only similarity measures that take into account a certain group of drugs compared to another group of drugs might not work in such case. Such similarities are desirable if one intends to model the characteristic drug features for a given target family. As shown in Figure 4.4, drugs for various protein targets do not show similarities to drugs for other targets even if they are targeting proteins involved in NDD disorders. As the similarity measures do not provide a better solution to discriminate NDD drugs from non-NDD drugs, we investigate the descriptor space for NDD drugs to evaluate whether there are shared features among them, to identify them as NDD drugs. Therefore, we studied all the NDD drugs as a single data set regardless of their respective target protein families on their available molecular descriptors side.

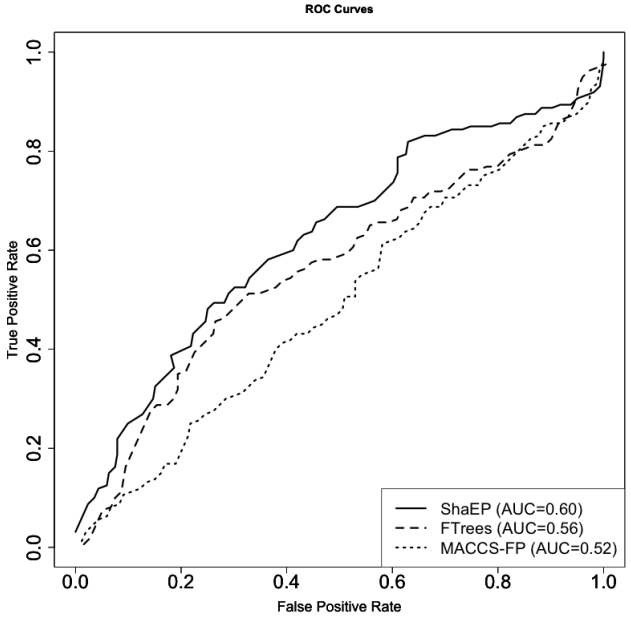


FIGURE 4.3: ROC curves for three different similarity measures.

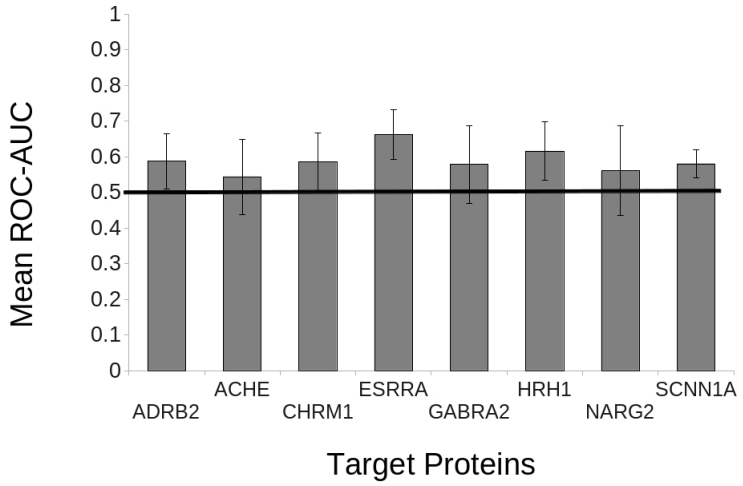


FIGURE 4.4: Mean ROC-AUC values for similarity comparisons of individual target family specific drugs with the rest of the NDD drugs. The central line indicates a random performance. Descriptions of target proteins are given in Table 4.3.

TABLE 4.3: Full names and gene symbols of the target proteins for which the similarity comparisons of their interacting drugs are performed using different similarity measures.

No	Gene Symbol	Full name
1	ADRB2	beta-2 Adrenergic Receptor
2	ACHE	Acetylcholinesterase
3	CHRM1	Muscarinic Acetylcholine Receptor M1
4	ESRRA	Estrogen-related Receptor alpha
5	GABRA2	gamma-Aminobutyric Acid (GABA) A Receptor, alpha 2
6	HRH1	Histamine Receptor H1
7	NARG2	NMDA Receptor-regulated Protein 2
8	SCNN1A	Sodium Channel, non-voltage-gated 1 alpha subunit

4.3.3 Comparison with LDA Classification

The LDA based method was tested as a comparison to SVM-RFE for the binary classification of NDD drugs from non-NDD drugs. The LDA based feature selection and classification algorithm generated a set of ranking descriptors that can be used in the binary classification, in a similar fashion as SMV-RFE. However, the classification performance obtained with SVM-RFE was better than that obtained

with LDA for all three sets of molecular descriptors. The comparisons are given in Table 4.1. An overall accuracy of 0.72 was obtained for the independent test set using LDA based ranked descriptors while accuracy of 0.80 was achieved by using SVM-RFE based ranked descriptors. In our experimental evaluation, however, DDA case of LDA performs poorly for understandable reasons since it does not assume any correlation among features. LDA with up to 72% accuracy performs better than DDA (that has accuracy below 60%), which is still outperformed by SVM-RFE (see Table 4.1). We assume that this is due to the inherent power of non-linear kernel machines (SVM) that is achieved by projecting data in high dimensional non-linear space.

4.3.4 Application in NDD Focused Library Design

The predictions made by using the top ranking descriptors can also be ranked based on the decision values or probabilities calculated by the underlying machine learning model. This is very useful in predicting drug molecules with ranges of probability that presents the likelihood that the molecule will belong to a certain class. We rank the predictions made by the machine-learning model based on the decision values or probabilities that are assigned for each drug that falls into a certain class of drugs. Based on the decision values (Eq. 4.1) as shown in Figure 4.5, ranks for the corresponding predictions can be used to filter a focused library of drug-like molecules from large chemical libraries. Figure 4.5 shows that values above the line at 0 are predictions for drugs as NDD+ while values below 0 are predictions for drugs as NDD-. Clusters of data points falling at -1 or 1 (as seen in Figure 4.5) for most of the drugs represent that these drugs have similar distances from the support vectors and have been assigned similar decision values based on the selected descriptor dimension space. Applying the same model to the data of a large compound library for which only the top ranked descriptors are calculated would classify them with a range of probabilities. These probabilities and rankings can be used here to design a focused library by considering the predictions with higher probabilities of falling in a certain class of interest with good confidence level. It remains, however, to be shown that this assumption holds true empirically if the model learned is applied to physical focused library construction.

4.3.5 Pharmacological Modeling of Approved Drugs

In order to test the applicability of the proposed model for pharmacological modeling of drugs, the SVM model was applied to a set of all approved drugs collected

from DrugBank. The top ranked 40 descriptors were calculated for these drugs and the drugs, which were already present in the training set, were excluded in order to avoid artificial bias. The new predictions were sorted using the ranks calculated by the SVM model and top 50 predicted drugs (i.e 5% of all approved drugs) were selected as shown in Figure 4.5. The predicted drugs were further analyzed by calculating their individual structural similarities with NDD drugs and their interacting targets were collected from DrugBank, STITCH DB and CTD databases.

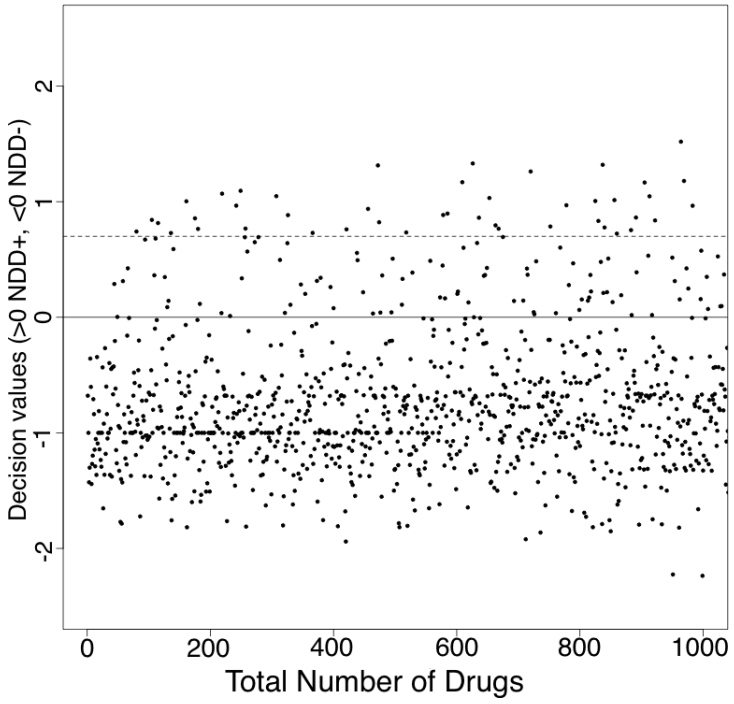


FIGURE 4.5: Plot of decision values assigned to the approved drugs data set by the SVM model. The points above the horizontal line at 0 indicate the drugs predicted as NDD drugs. The dotted line indicates a threshold of selecting the top 5% predicted drugs for analysis.

Based on the similarities to known NDD drugs (Tanimoto coefficient ≥ 0.6) and having interactions with common drug targets, 35 drugs were identified. 12 out of these 35 drugs have known anticonvulsant properties. Further more, protein binding site similarities were retrieved from in-house database of pre-computed protein pockets similarities. This database contains the 3D structures of all target proteins interacting with approved NDD drugs and their binding site similarities values. Pair-wise similarities between binding pockets are calculated using the SMAP software package[94]. SMAP detects the binding sites in protein structures and performs similarity comparison of protein 3D structures independently of their

sequence. Binding site similarities are computed for more than 2000 proteins with PDB structures and significant similarities are stored in the database.
The binding site similarities are explored in order to observe the pocket geometries of multiple proteins that are targeted by a group of similar drugs. The selected drugs were arranged in a network of drugs interacting with target proteins for visualizations as shown in Figure 4.6 and Figure 4.7. The drugs nodes are represented by rectangles and the protein targets are shown as circles. An undirected edge between two drugs indicates structural similarity and that between two proteins indicates binding site similarity. Likewise, an edge between a drug and a protein indicates the interaction of that drug with its protein target.
Interestingly, the structural similarities of the predicted drugs with NDD drugs are not so high. However, the predicted drugs either share protein targets directly or indirectly through interacting with an NDD drug targeted protein that has a similar binding pocket. This implies that the prediction of new drug-target associations of the predicted drugs can be used to streamline drug repurposing [128] and may result in novel insights that can lead to the identification of novel drug actions in a different indication area.

4.4 Discussion

A better understanding of the mechanisms by which drugs act against complex diseases such as neurodegenerative disorders can be achieved by integrating all available drug-target interaction data at systems pharmacology level to derive *in-silico* computational models. We have collected data for drugs that act against neurodegenerative disorders from various databases. The drug data is analyzed for their specific molecular features that confer them with approved efficacy and safety profile to such drugs. These drugs are active against a variety of protein targets belonging to different protein families and different pathways. We have observed that similarity measures based on a set of target specific drugs, do not produce good results in terms of similarity predictions when used in queries. On the other hand, using machine-learning approaches to study the drugs collectively for their features offer an alternative to conventional similarity-based approaches. There are hundreds of molecular features that characterize various physiochemical properties of the molecules. However, a number of molecular descriptors will measure the molecular properties in a similar way. Therefore, correlations among these features should be considered. As shown in our results, removing strongly correlated features reduced the total number of features by one-third in the first

TABLE 4.4: List of predicted and known NDD drugs with their indication areas and structural similarities. Indication areas for similar NDD drugs with references are given in Table 4.5

Predicted Drug	Indication	Similar NDD Drug	Indication	Similarity
Lamivudine	HIV-1; HBV	Cladribine	Multiple Sclerosis	0.67
Pentazocine	Pain	Apomorphine	Parkinsons Disease	0.63
Propantheline	Antispasmodic; rhinitis	Diphenhydramine	Parkinsons Disease	0.65
Dronabinol	Antipsychotic	Vitamin E	Alzheimers Disease	0.80
Mestranol	Contraception	Estriol	Multiple Sclerosis	0.80
Desogestrel	Contraception	Prednisone	Multiple Sclerosis	0.62
Levallorphan	Drug overdose	Naltrexone	Seizures	0.68
Sulfamethizole	Antibacterial	Acetazolamide	Epilepsy	0.88
Sulfanilamide	Vulvovaginitis	Dapsone	Seizures	0.80
Glutethimide	Insomnia	Ethosuximide	Epilepsy	0.71
Meprobamate	Anxiety; muscle relaxant	Felbamate	Epilepsy	0.67

place. Furthermore, there are relevant molecular descriptors that should be properly selected for efficiently discriminating the molecular data. Selection of the most relevant descriptors is achieved by recursively eliminating the less relevant descriptors.

Using an SVM-based machine learning approach, the most relevant descriptors can be selected and used to perform predictions on a larger set of molecular data. The generalization errors obtained by selecting the varying number of top ranking descriptors inform us about the model's classification error that can occur when a given number of top ranking descriptors is used to classify the data. This is very beneficial for reducing the high dimensionality of the drugs descriptors space and selecting a minimum number of descriptors that can be calculated efficiently for a large number of molecules, hence avoiding compute intensive and time consuming calculations. Therefore, with the help of this algorithm, a limited number of top ranking descriptors can be calculated to enrich a molecular library or design an NDD focused library of molecules by considering the probabilities assigned during classification.

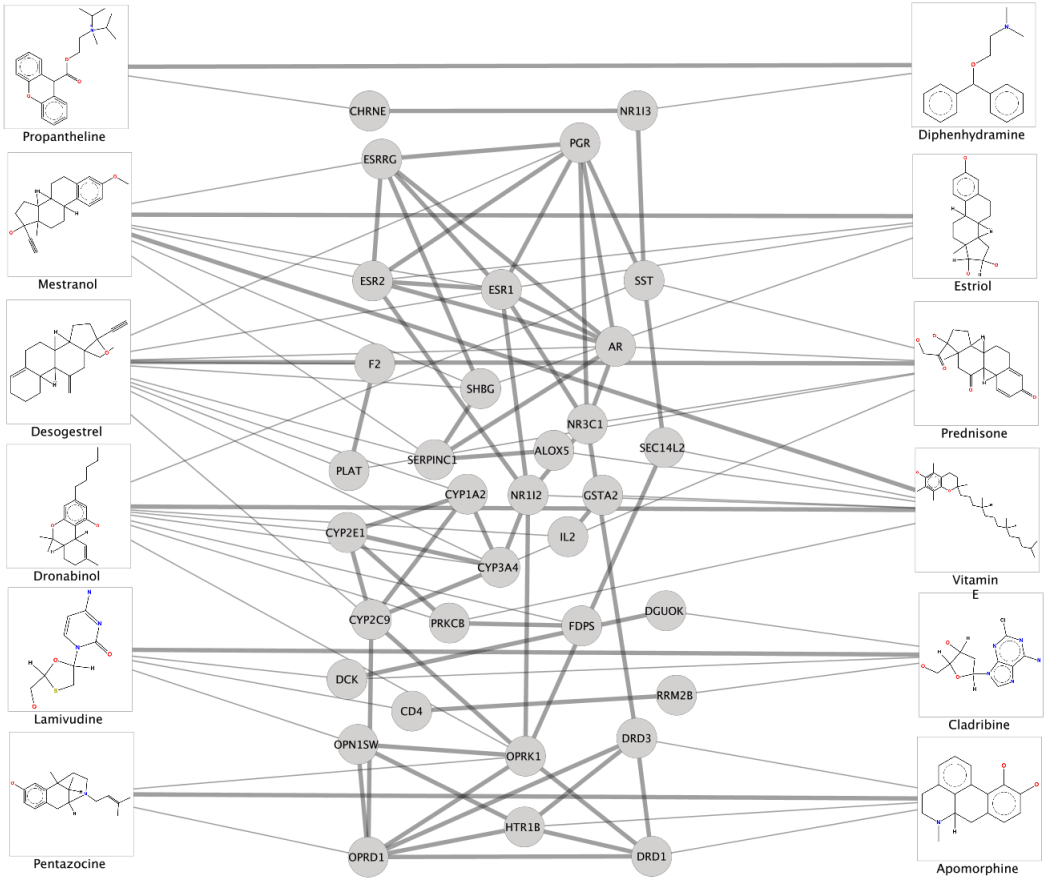


FIGURE 4.6: Drug target network of the selected predicted drugs (left) and their similar NDD drugs (right) with their common interacting protein targets (center). Thick lines indicate similarities between drugs and binding site similarities between drug-targets. Thin lines indicate interaction of drugs with the target genes.

The high sensitivity of the presented SVM-based model to classify NDD drugs from non-NDD drugs and its satisfactory accuracy of discriminating NDD drugs from non-NDD drugs is in agreement with previous studies that report on the efficiency of SVM approach to predicting properties of chemical compounds and classifying drug-like compounds from others [115], [116]. In fact, identification of the most important structural features and chemical properties of successful drugs is the first step towards addressing the challenge of design and optimization of next-generation multi-target drugs.

4.5 Conclusions

Making the systems pharmacology paradigm a reality requires designing multi-target compounds with optimized structure-activity relationship (SAR) in such a

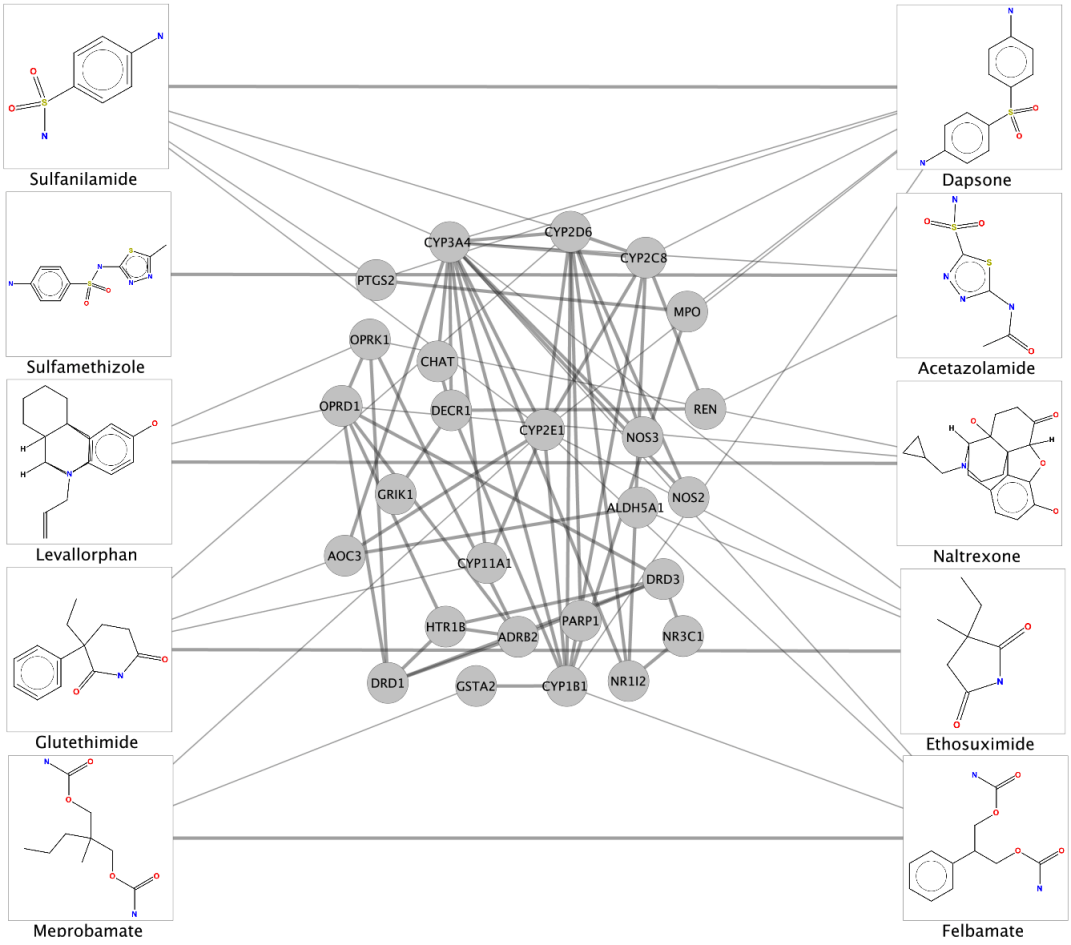


FIGURE 4.7: Drug target network of some predicted drugs (left) and their similar anti-epileptic drugs (right) with their common interacting protein targets (center). Thick lines indicate similarities between drugs and binding site similarities between drug-targets. Thin lines indicate interaction of drugs with the target genes.

way that their pharmacological profile towards selective multiple targets is maintained. Computational approaches, as we have demonstrated here with the use of machine learning methods, can play a major part in identifying molecular descriptors that confer certain pharmacological properties to candidate compounds. Statistical and machine learning methods have previously been used to predict permeability of CNS drugs. Here we used the learning methods to predict compounds that can be used in NDD indications. We also demonstrated the advantage of such methods over the conventional similarity based methods. A limitation in the development of such classifier models is the lack of enough data for sampling and training. Apart from that, the available data in such models is not experimentally verified to be true negative, therefore, the negative data is not often confirmed negative which can ultimately result in the performance loss of the underlying classifier models. This method can be effectively applied to design a highly enriched

focused library of compounds from a large compound database. The proposed method has applicability in drug repurposing and some drugs that are used in non-NDD indications are predicted to be potential candidates for their usefulness in various CNS disorders. Obviously such predictions, to turn out successful, must be coupled with modeling and experimental validation for target combinations.

TABLE 4.5: List of known NDD drugs with their indication areas from DrugBank and CTD/TTD databases.

NDD Drug	Indications (DrugBank Database)	Indications (CTD/TTD Database)
Cladribine	For the treatment of active hairy cell leukemia (leukemic reticuloendotheliosis) as defined by clinically significant anemia, neutropenia, thrombocytopenia, or disease-related symptoms.	Therapeutic role in Multiple Sclerosis (1 reference in CTD database)
Apomorphine	For the acute, intermittent treatment of hypomobility, off episodes (end-of-dose wearing off and unpredictable on/off episodes) associated with advanced Parkinson's disease.	Therapeutic role in Seizures, Parkisons Disease (23,16 references respectively in CTD database)
Diphenhydramine	For the treatment of symptoms associated with Vertigo/Meniere's disease, nausea and vomiting, motion sickness and insect bite	Therapeutic role in Seizures, Parkinsons Disease (14, 1 references respectively in CTD database)
Vitamin E	Vitamin E, known for its antioxidant activities, is protective against cardiovascular disease and some forms of cancer and has also demonstrated immune-enhancing effects. It may be helpful in some neurological diseases including Alzheimer's.	Therapeutic role in Neurodegenerative Diseases, Seizure (10,19 references respectively in CTD database)
Estriol	Used as a test to determine the general health of an unborn fetus.	Multiple Sclerosis (1 reference in TTD database.
Prednisone	For the treatment of drug-induced allergic reactions, multiple sclerosis and many other disease	Therapeutic role in Multiple Sclerosis (3 references in CTD database)
Naltrexone	Used as an adjunct to a medically supervised behavior modification program in the maintenance of opiate cessation in individuals who were formerly physically dependent on opiates and who have successfully undergone detoxification.	Therapeutic role in Seizure (14 references in CTD database).
Acetazolamide	For adjunctive treatment of: edema due to congestive heart failure; drug-induced edema; centrencephalic epilepsies; chronic simple (open-angle) glaucoma.	Therapeutic role in Seizure, Epilepsy (3 references against each in CTD database).
Dapsone	For the treatment and management of leprosy and dermatitis herpetiformis.	Marker mechanism in Seizure (6 references in CTD database).
Ethosuximide	For the treatment of petit mal epilepsy.	Therapeutic role in Epilepsy (11 references in CTD database).
Felbamate	For use only in those patients who respond inadequately to alternative treatments and whose epilepsy is so severe that a substantial risk of aplastic anemia and/or liver failure is deemed acceptable in light of the benefits conferred by its use.	Therapeutic role in Epilepsy (10 references in CTD database).

Chapter 5

SVM based Classification of GPCR Antagonists and its Application in Drug Repurposing

5.1 Introduction

G-Protein Coupled Receptors (GPCRs) play important roles in the healthy human physiology and they represent the largest family of drug targets [129]. They are the targets of 30 to 50% of all marketed drugs [130] that exhibit their pharmacological actions through GPCR-associated mechanisms. GPCRs are among the most prominent proteins in the receptors family, hydrolases in the enzyme family and voltage-gated Ca channels in the ion-channel family [131]. Gene family distribution of targets and drugs revealed that more than 50% of drugs target only four gene families [132] as shown in Figure 5.1:

a Class I GPCRs

b Nuclear Receptors

c Ligand-gated Ion Channels

d Voltage-gated Ion Channels

5.1.1 GPCRs as Attractive Therapeutic Targets

GPCRs contribute a significant number of therapeutic targets to the current number of drug targets that are in the order of 10^2. The discovery of new drugs acting against multiple GPCRs has been successful for decades and account for

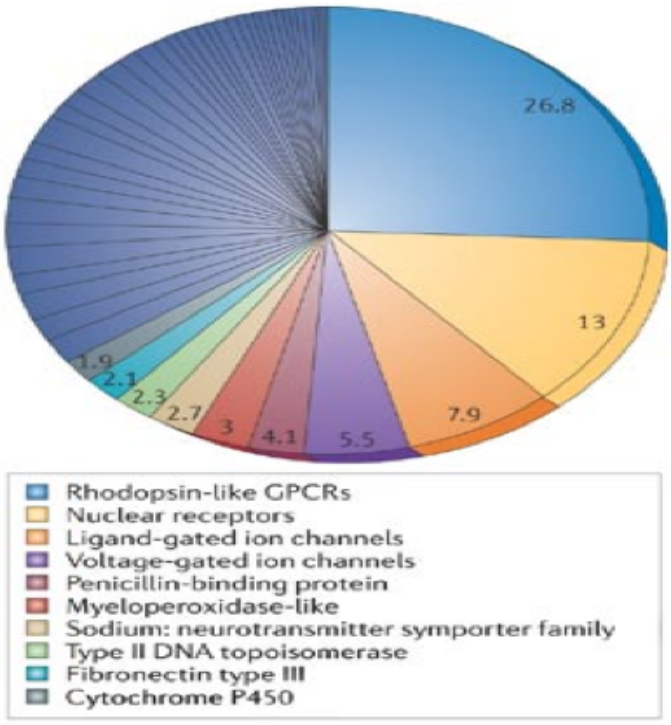


FIGURE 5.1: The percentage of all FDA-approved drugs based on the top 10 gene familes is displayed [132].

approximately 50% of all launched drugs [133]. Similarly, on the drug targets side, GPCRs represent 50-60% of current drug targets for their involvement in different indications such cardiovascular, metabolic, neurodegenerative, psychiatric and oncologic diseases [134]. They can accommodate thousands of different ligands (e.g. agonists, antagonists or inverse agonists) because of their common binding site structures bundled through seven-transmembrane α-helices [135]. G protein-coupled receptors play an important role in many physiological functions and pathophysiological processes involved in many diseases, including cancer and cancer metastasis [136]. They are the subjects of intense research in both academia and pharmaceutical industry [137]. Due to their important mediation, functions, the GPCRs family of targets are one of the most successful targets for a broad spectrum of diseases [138]. There is a continuous attempt to "de-orphanize" the GPCRs among the large receptorome (the portion of the proteome encoding receptors) to better characterize receptors with unknown functions [139]. Thus a better understanding of signaling and regulatory mechanisms of GPCRs should open new perspectives for GPCR-directed drug discovery [140].

5.1.2 Role of GPCRs in Neurodegenerative Disorders

The role of GPCR trafficking and cell membrane expression is very important in the abnormal processes during diseases caused by receptor misfolding [141] and GPCRs are subjected to cellular quality control system. They play pivotal role in regulating the function and plasticity of neuronal circuits in the central nervous system due to their involvement in G-protein signaling pathway [142]. Alterations in these signaling pathways have been extensively investigated in various types of neurodegenerative diseases and neuro-inflammatory disorders [143]. Various studies have investigated the role of GPCRs in Parkinson's disease, Huntington's disease [144], [145], Alzheimer's disease [146], Amyotrophic Lateral sclerosis and Multiple Sclerosis [147],[148],[149]. Current Alzheimers disease therapeutics mainly target Acetyle-choline esterase that stimulates cholinergic neurons [150], even though the complex neurodegenerative process is not limited to a specific neurotransimitter system [151]. Several studies have been reported the critical role of GPCRs in the pathogenesis and amyloidogenesis of Alzheimers disease [147]. Apart from the neuroprotection therapy through inhibition of mono-amine oxidase-B [152], disruption in GPCR signaling is also a disease mechanism in Parkinson's disease [153], [154]. Similarly, GPCR such as Histamine receptors play a potential role as therapeutic targets for anti-epileptic drugs [155]. Heteromerization of different GPCRs can have its effect on the activation of a variety of signaling pathways in neurological disorders [156]. For example, functional interactions between adenosine and dopamine receptors in basal ganglia [157] and effect of adenosine receptors on glial functions [158] highlight the potential of developing antagonists for these receptors to control neuromodulation in aspects relevant to Parkinson's disease.

5.1.3 GPCR Directed Ligand Library Design

The recent availability of a few high-resolution GPCRs structural data has opened new opportunities for computational modeling to improve our understanding about them [159]. Comparative homology modeling of 3D structures of GPCRs is a tool that allows understanding of receptor-ligand interactions [160] and because of the limited availability of GPCR crystal structures, theoretical coordinates can be made available [161]. However, such homology based models require validations [162]. With little and slowly expanding structural information of GPCRs targets, validation of computational modeling approaches crucially depends on experimental data [137]. Different chemogenomic approaches, such as ligand centered, protein centered and protein-ligand centered [163], can be applied to the design and

exploration of GPCR ligand space. The ligand-based approach relies on physiochemical properties of GPCR ligands [164] which can be used to categorize ligands into "GPCR-ligand-like" and "non-GPCR-ligand-like" [165]. The structure-based design methods rely on the three dimensional structure of proteins for rational drug discovery [166] and can lead towards a better understanding of GPCR activation [167]. Once the focused libraries are designed, they can be combined with structure-based methods, which also take into account, the structural information and biological knowledge of the target receptors [168]. There exist several specialized GPCR databases that collect information about receptors (e.g. GPCRDB [169], its natural variants in GPCR NaVa [170]) and GPCR ligand databases (e.g. GLIDA [171] and GLL/GDD [172]). These databases provide useful information about the GPCR target families and their respective ligands that we can use for the design and validation of our *in-silico* models.
This chapter focuses on our study of GPCR antagonist ligands that includes:

a The exploration of molecular and physiochemical properties (i.e. descriptors/features) of GPCR antagonist ligands and evaluate their usefulness in discrimination of GPCR related bioactive molecules from the inactive molecules.

b The application of machine learning based recursive feature elimination method [117] to select and compare sets of molecular descriptors for the classification of GPCR antagonists.

c The building of molecular descriptors based classifier models that will be used to classify unknown ligands and GPCR subfamily ligands

d The application of such classifier models in the repositioning of approved drugs that are used for other indications.

5.2 Methods and Material

5.2.1 Data Sets

A data set of antagonist ligands of four GPCRs is compiled from the GLL/GDD database [172]. The GLL/GDD database contains more than 25,000 agonists and antagonists ligands for 147 GPCR targets and approximately 1 million decoy compounds, with a ratio of 39 decoy compounds per ligand. We preferentially selected antagonist ligands over the agonist ligands for building the classifier models because there are more antagonist than agonist in this database for our selected

GPCR receptors, and we are interested in designing antagonists for these receptors later on. Four human GPCR receptors (beta 2 adrenergic receptor (ADRB2), adenosine A2A receptor (ADORA2A), dopamine D3 receptor (DRD3) and histamine H1 receptor(HRH1)) are selected due to their suggested polypharmacological profile that arise from their binding site similarities as described in more detail in the chapter "Drug Target Networks Analysis". The distribution of their antagonist ligands and decoys is listed in Table 5.1, which comprise total of 960 antagonists and 4000 decoy molecules for the selected targets. For the reason of simplicity, for the fast calculation of molecular descriptors and to avoid bias that can be introduced by including a large number of decoy compounds [173], we selected 1000 decoy compounds per target.

TABLE 5.1: Total number of antagonist ligands and decoy compounds per GPCR target.

Target	**Antagonists**	**Agonists**	**Decoys**
Adenosine A2A Receptor	361	82	1000
beta-2 Adrenergic Receptor	204	206	1000
Dopamine D3 Receptor	317	6	1000
Histamine H1 Receptor	78	8	1000

5.2.2 Descriptors Calculation

A set of all 2D and 3D QSAR descriptors (features) are calculated for the antagonists and decoy compounds using MOE [92]. The compounds structures were energy minimized with MMFF94x [119] force field within MOE prior to calculation of descriptors. The resulting descriptors that are meaningful were kept, while some meaningless descriptors that have a value of 0 or 1 for all compounds, were removed. Values of some descriptors could be zero (e.g. "violate lipinski" rule will result in 0 values for all the ligands that do not violate this rule) and value for a descriptor could be 1 for most compounds (e.g. "is drug like"). This resulted in a total number of 274 molecular descriptors out of the total 334 available descriptors. Similarly, compounds with missing values for some descriptors, which are not calculated for various reasons, are also removed from the data sets. The data set is randomly divided into a training set (474 antagonists, 2000 decoys) on which an SVM model is trained and a test set (472 antagonists, 2000 decoys) on which the model is evaluated.

5.2.3 Descriptors Selection/Ranking

A ranked list of molecular descriptors is obtained by using the machine-learning algorithm with recursive feature elimination (RFE) as discussed in the previous SVM based Classification of NDD Drugs for Pharmacological Modeling chapter [39]. This algorithm ranks features according to their contribution in binary classification of the GPCR antagonists from the decoys compounds. A list of top-ranked molecular descriptors is given in Table 5.2. These selected top-ranked molecular descriptors are used to build and train the SVM model on the training data set.

Table 5.2: Top ranked molecular descriptors selected by SVM-RFE

No.	Code	Descriptor	Average Rank
1	E_strain	Local strain energy: the current energy minus the value of the energy at a near local minimum.	1.0
2	E_vdw	van der Waals component of the potential energy.	4.6
3	E_ele	Electrostatic component of the potential energy.	8.5
4	pmiZ	z component of the principal moment of inertia (external coordinates).	8.9
5	E_nb	Value of the potential energy with all bonded terms disabled.	9.4
6	E_str	Bond stretch potential energy.	9.8
7	pmiX	x component of the principal moment of inertia.	18.5
8	E_oop	Out-of-plane potential energy.	21.0
9	SlogP_VSA0	The Subdivided Surface Areas are descriptors based on an approximate accessible van der Waals surface area (in $Å^2$) calculation for each atom, v_i along with some other atomic property, p_i	21.2
10	GCUT_PEOE0	The GCUT descriptors are calculated from the eigenvalues of a modified graph distance adjacency matrix.	21.5

Interestingly, the top ranked descriptors as shown in Table 5.2 are energy and electrostatic based molecular descriptors, which is in strong agreement to the charge related molecular properties of GPCR ligands. The vast majority of ligands for several GPCR targets have non-zero charge [172], that contributes towards the distinguished features of these GPCRs antagonists ligands.

5.3 Results and Discussion

Classification accuracies of ∼0.98 have been observed for 10x internal cross validation and independent test sets. These results show that the top ranked selected descriptors as given in Table 5.2 can be used to classify the data sets with a very good classification performance. These descriptors are assigned high average ranks among all 274 calculated that are calculated by RFE during internal 50 fold cross validation. The training data set is used to first select the top ranking descriptors from the 274 descriptors and then, on the basis of these few but highly relevant descriptors, the underlying data is classified into their respective GPCR antagonists and decoy classes. The overall accuracy on the training data set with a precision of 96% is shown in Figure 5.2. The classification result over the test set is shown in Figure 5.3 and shows that a similar performance is obtained for the binary classification.

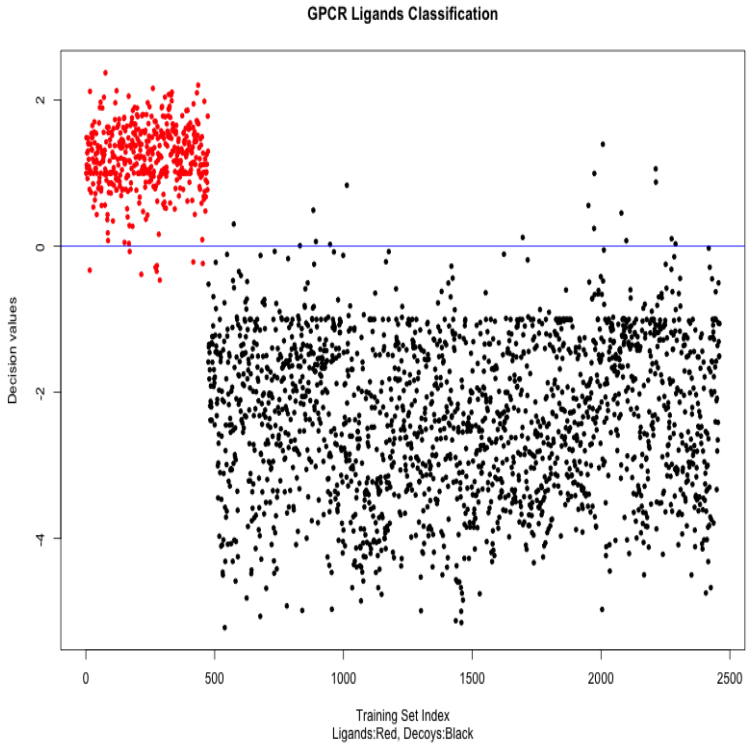


FIGURE 5.2: A decision values scatter plot involving the training data set. The points in red color represent the antagonist ligands while those in black color represent the decoy compounds. Decision values above 0 mean the data is classified in to GPCR antagonist class and below 0 mean the data is classified into the decoys class.

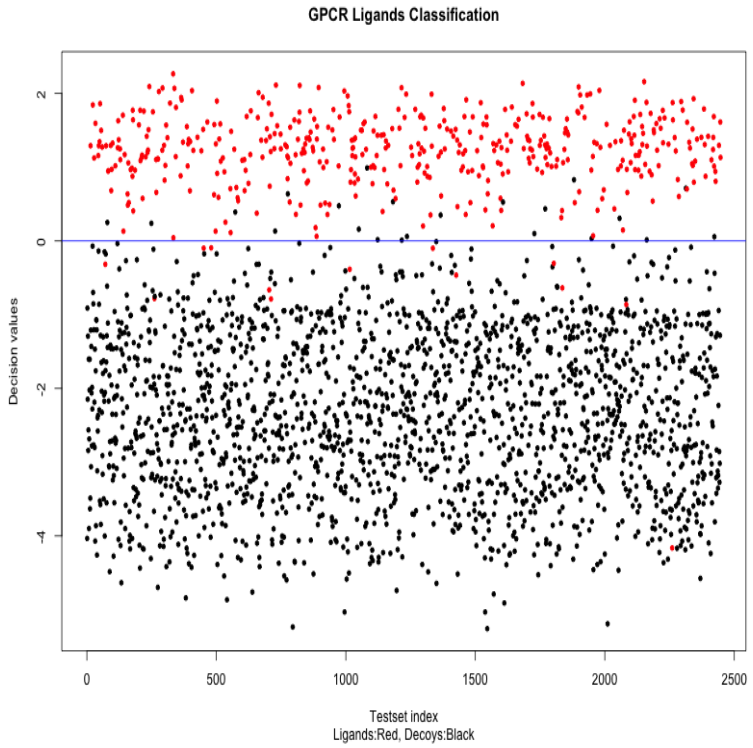


FIGURE 5.3: A decision values scatter plot involving test data set. The points in red color represent the antagonist ligands while those in black present the decoy compounds. Decision values above 0 mean the data is classified in to GPCR antagonist class and below 0 mean the data is classified into the decoys class.

5.3.1 Further Validation

The SVM model that is trained over the antagonist data of only four GPCR targets (Table 5.1) is further evaluated on the antagonist data for another set of 81 GPCR targets. Table 5.3 shows a list of some GPCR families and the total number of antagonists present in each family. A total of ∼18,000 antagonists are collected for the 81 GPCR targets and approximately 4,000 decoy compounds are randomly selected from the decoy sets of each target. The top 10 molecular descriptors are calculated for the antagonist and decoy compounds. Not all the decoys are used for the descriptors calculation because calculating descriptors for the decoys for all 81 GPCR targets account for more than 600,000 compounds, that would require more computation time. The SVM model is applied to predict the correct classes of antagonists and decoys and a resulting classification accuracy of 0.96 is achieved. However, there is about 15% overlap between the antagonists that are shared by some GPCR family members. When these duplicate ligands were removed, it did not affect the classification accuracy. Classification results

are depicted in Figure 5.4 which shows that antagonists are fairly discriminated from the decoy compounds.

TABLE 5.3: List of some GPCR target families and sub-types and total number of antagonists per family.

Family	**Name**	**Subtypes**	**Antagonists**
5HT	5-hydroxytryptamine	1/2 A,B,C,D,E,F	2436
AA	Adenosine	1,2,3, A,B	1106
ACM	Muscarinic acetylcholine	1,2,3,4,5	586
ADA	Alpha adrenergic	1/2 A,B,C,D	3016
ADR	Beta adrenergic	1,2,3	585
DRD	Dopamine receptors	1,2,3,4,5	1711
HRH	Histamine receptors	1,2,3,4	529
OPR	Opioid receptors	D, K, M, X	75
PE2R	Prostaglandin E2 receptors	1,2,3,4	512
NK	Substance-K/P receptors	1,2,3	1164
Total			11720
Other	other families		6585
Duplicates			2616
Decoys			4,000

5.3.2 Multiple Classification

Multiple SVM classifier is evaluated for the classification of multiple classes of GPCR antagonists at their target family level. Multiple classification, contrary to binary classification that only classifies GPCR-ligands and non-GPCR-ligands, was used to distinguish between antagonist ligands that are specific for different GPCR families. A total of 6533 antagonists from the GLL database are grouped into 8 target families (as listed in Table 5.4 with respective number of antagonists per target family) and then the descriptors based multiple classification is evaluated. All 274 descriptors were used for the multiple classification that were previously calculated for the binary classification. Approximately 50% of agonists from each target family were used in the training set and the rest were kept in the test set. Duplicate antagonists were removed from both training and test sets. MSVMpack [174], an open source package for multi-class support vector machines was used for the classification of antagonists of multiple GPCR target families.
An overall classification accuracy of 86.7% on the test data set is obtained for classifying antagonists of individual target families. Classification matrix (also known as confusion matrix) over the test data set has been given in Table 5.5. Values across diagonal of the matrix are the number of antagonist samples correctly

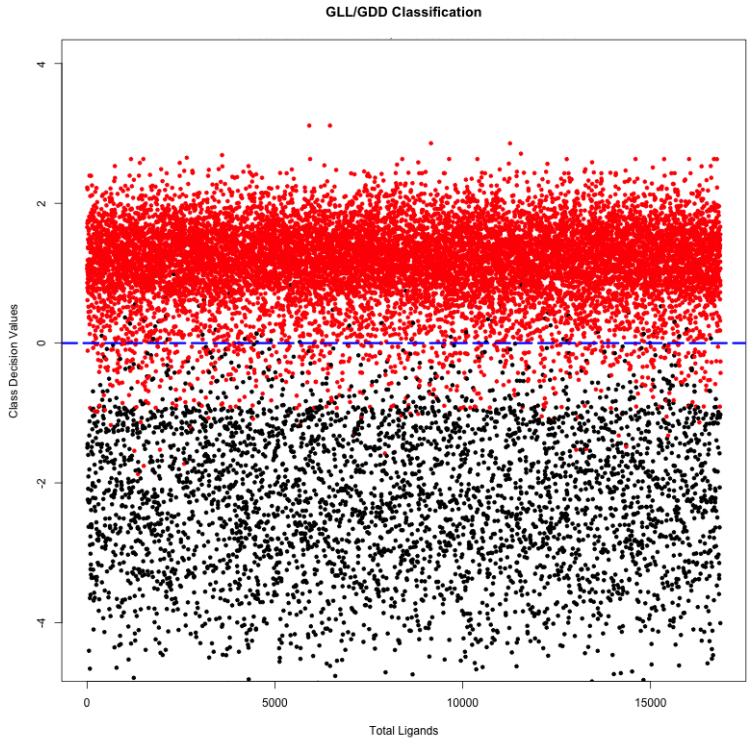


FIGURE 5.4: Classification of antagonists/decoys for 81 GPCR targets from GLL/GDD database. GLL antagonists are represented in red color and GDD decoy compounds are represented in black color. The decision boundary line is drawn horizontally at 0.

TABLE 5.4: List of GPCR target families grouped into classes with corresponding number of antagonists and decoys per target family class used in multiple classificaiton.

Class	**Family**	**Subtypes**	**Antagonists**	**Training set**	**Test set**
1	5HT	1/2 A,B,C,DE,F	1930	975	921
2	AA	A1, A2, A2B, A3	652	352	296
3	ADA	1A,B,D, 2A,B,C	780	386	373
4	ADRB	1,2,3	205	99	93
5	DRD	1,2,3,4,5	1300	649	587
6	HRH	1,2,3,4	474	95	83
7	NK	1,2,3	1053	544	494
8	PE2R	1,2,3,4	139	99	31
9	Decoys		5968	1000	4968

classified in to their respective target family classes, while the off-diagonal values denote the number of antagonists of one family class classified inside another family class. For example, 54 antagonists of ADRB were correctly classified into its ADRB class, 17 were misclassified into 5HT class, 1 was misclassified into AA

class, 6 were misclassified into ADA and so on. These results show that using molecular descriptors, the antagonists can be distinguished between the major GPCR families of targets. Such multiple classification scheme can help design target family specific libraries of compounds. In comparison to binary classification where we used the list of top-ranked descriptors produced by SVM-RFE, we used all of the descriptors in multiple classification as MSVMpack does not provide functionality for selecting and ranking descriptors. Although there is a number of antagonists shared by different target families, due to which the chances of misclassification would increase, MSVMpack still performed good predictions of the classes for most of the target families.

TABLE 5.5: Confusion matrix of the multiple classification predictions over test data set for 8 GPCR target classes

	5HT	**AA**	**ADA**	**ADRB**	**DRD**	**HRH**	**NK**	**PE2R**	**Decoy**
5HT	703	23	47	1	131	0	15	1	0
AA	23	218	5	0	14	0	7	0	29
ADA	185	7	109	3	46	2	18	3	0
ADRB	17	1	6	54	2	0	2	0	11
DRD	167	0	21	0	380	0	5	0	14
HRH	46	0	2	0	21	14	0	0	0
NK	29	20	6	0	6	0	430	3	0
PE2R	0	2	0	0	0	0	7	22	0
Decoys	5	10	12	11	45	2	7	0	4876

5.3.3 Drug Re-purposing Application

The resulting performance achieved with SVM to classify antagonists from non-antagonists of GPCRs suggests that the classifier model can be applied on a set of drugs or compounds to predict their likelihood of being antagonist for GPCRs in general. In order to apply the classifier model on a data set of approved drugs, approximately 1400 FDA approved drugs were downloaded from DrugBank [**?**]. Molecular descriptors were calculated for the drugs as done before and the SVM classifier model was applied to them. The classifier model predicted approximately 60% of all approved drugs as GPCR binders as shown in Figure 5.5.

Ranking order over the predictions of SVM classifier model was performed on the basis of the probability values for each drug obtained by SVM. The top 20 high ranking predictions were selected for subsequent analysis. For these selected drugs,

their interacting targets and mechanism of action information were also retrieved from DrugBank. It was observed that 9 out of the 20 predicted drugs had a GPCR listed as a target and 5 drugs are targeting bacterial proteins as shown in Table 5.6. The 2D structural diagrams of these drugs are depicted in Figure 5.6. These drugs made a set of potential candidate drugs that can be used for drug re-purposing study, particularly the antibiotic drugs can be investigated for their potential use in other indication areas. The drugs that are already interacting with GPCR targets mentioned in DrugBank can also be studied for their interactions with other target families of GPCRs that might suggest new indication areas for these drugs.

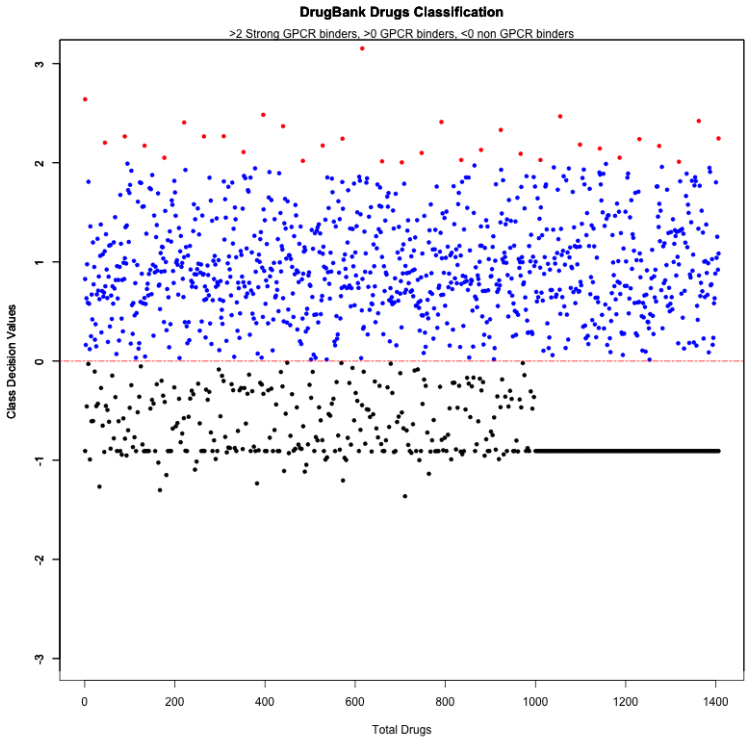


FIGURE 5.5: Classification results obtained for all FDA approved drugs from DrugBank. The data points above the decision boundary line represent the drugs classified as active against GPCRs. The drugs presented with red color are suggested as the drugs with higher probability of being active against GPCRs.

TABLE 5.6: List of FDA approved drugs, their targets and indications from DrugBank, which are predicted as GPCR binders.

DrugBank ID	**Targets**	**Indication**
DB00769	NR3C1, CYP3A4	Used topically as an anti-inflammatory in the treatment of steroid-responsive dermatoses.

DB00177	AGTR1, CYP2C9	Used as a first line agent to treat uncomplicated hypertension, isolated systolic hypertension and left ventricular hypertrophy.
DB00485	LMHCC_2184, LMHCC_2773, pbp1b, bp2a, pbp3,pbpA, penA	Used to treat infections caused by penicillinase-producing staphylococci that have demonstrated susceptibility to the drug.
DB01190	rplJ, 23S rRNA, CYP3A4	For the treatment of serious infections caused by susceptible anaerobic bacteria, including Bacteroides spp., Peptostreptococcus, anaerobic streptococci, Clostridium spp., and microaerophilic streptococci.
DB06718	Stanozolol Binding Protein (STBP)	Used in treating C1-inhibitor deficient hereditary angioedema.
DB00947	ESR1, CYP3A4	For the treatment of hormone receptor positive metastatic breast cancer in postmenopausal women with disease progression following anti-estrogen therapy.
DB00301	pbpA, CYP3A4	Used to treat bacterial infection by susceptible microorganisms.
DB00496	CHRM1-5,CYP3A4, CYP2D6	For the treatment of overactive bladder with symptoms of urge urinary incontinence, urgency and frequency.
DB01046	CLCN2, CBR1	For the treatment of chronic idiopathic constipation in the adult population. Also used for the treatment of irritable bowel syndrome with constipation in women who are 18 years of age or older.
DB00332	CHRM1-3,CYP3A4, CYP2D6	For maintenance treatment of bronchospasm associated with chronic obstructive pulmonary disease, including chronic bronchitis and emphysema.
DB00219	CHRM1	For the treatment of visceral spasms.
DB00319	pbp3, penA, pbp2a, pbp1b	For the treatment of polymicrobial infections.
DB08795	pbp2a, pbp1b, pbp3, pbpA, penA	For the treatment of infection (Respiratory, GI, UTI and meningitis) due to E. coli, P. mirabilis, enterococci, Shigella, S. typhosa and other Salmonella, nonpenicillinase-producing N. gononhoeae, H. influenzae, staphylococci, streptococci.

DB00728	CHRNA2, CHRM2, HTR3A	Used as an adjunct to general anesthesia to facilitate both rapid sequence and routine tracheal intubation, and to provide skeletal muscle relaxation during surgery or mechanical ventilation.
DB01413	mrdA, ftsl, pbpB,pbpA	For the treatment of pneumonia (moderate to severe) caused by Streptococcus pneumoniae, including cases associated with concurrent bacteremia, Pseudomonas aeruginosa, Klebsiella pneumoniae.
DB00209	CHRM1, CYP2D6	For the treatment of overactive bladder with symptoms of urge urinary incontinence, urgency, and urinary frequency, detrusor instability and frequency of micturition.
DB01196	MAP2, MAP1A, ESR1,ESR2,CYP3A4	For the palliative treatment of patients with metastatic and/or progressive carcinoma of the prostate.
DB00704	OPRD1, OPRK1	Used as an adjunct to a medically supervised behavior modification therapy in the maintenance of opiate cessation in individuals who were formerly physically dependent on opiates and who have successfully undergone detoxification.
DB00246	DRDs, HTRs, ADRAs, CHRMs, CYP3A4	For the treatment of schizophrenia and related psychotic disorders.
DB01586	AKR1C2, CYP2E1	The drug reduces cholesterol absorption and is used to dissolve (cholesterol) gallstones in patients who want an alternative to surgery.

5.3.4 Targets Analaysis of Predicted GPCR Drugs

Targets information for the predicted drugs were also retrieved from the STITCH database [60] to see if they have any GPCR interacting target. Target information retrieved from STITCH database showed that there were GPCRs and other nuclear receptors that are interacting with 10 of the predicted drugs for which there was no GPCR target mentioned in the DrugBank as shown in Table 5.7. This tells us that the predictions made by the machine learning model are correct. Further *In-silico* validations are performed by comparative QSAR modeling and analysis tools available in MetaDrug [175] database.The selected drugs that are predicted

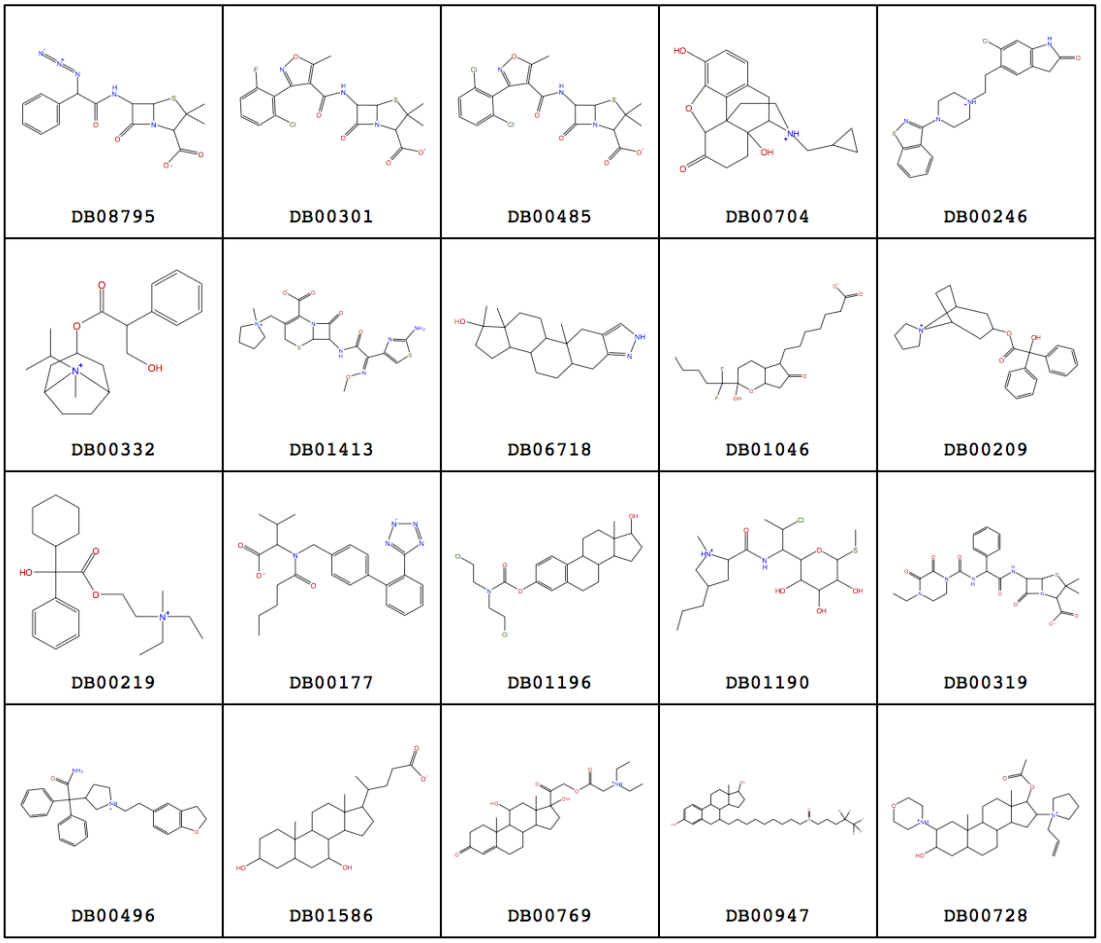


FIGURE 5.6: 2D structural representation of the top 20 approved drugs predicted as GPCR binders.

as potential candidate drugs for use against GPCRs are compared with a set of known GPCR drugs that are interacting with multiple GPCR targets. A set of 22 known GPCR drugs is selected from GLL database. *In-silico* comparison was made for the two sets of drugs by using QSAR based algorithms available in MetaDrug. The comparison results performed with MetaDrug prediction algorithms are shown in Figure 5.7 which shows that as the known GPCR drugs are interacting with multiple GPCR targets, most of the predicted drugs are also interacting with multiple GPCR targets.

5.3.5 Discovery of Mode of Action by Gene Expression Profiling

The mode of action (MoA) of some antibiotic drugs that are predicted as GPCR binders in this study, was analyzed by studying their similarity to drugs that have known transcriptional responses to drug treatments. The online tool MANTRA

TABLE 5.7: Interacting targets information retrieved for the top 20 drugs predicted as GPCR binders from DrugBank and STITCH Database.

Drug	**Targets in DrugBank**	**Targets in STITCH**
Hydrocortamate (DB00769)	NR3C1, CYP3A4	Nuclear harmone receptors
Valsartan (DB00177)	AGTR1, CYP2C9	
Dicloxacillin (DB00485)	LMHCC_2184, LMHCC_2773, pbp1b, bp2a, pbp3,pbpA, penA	GPR44
Clindamycin (DB01190)	rplJ, 23S rRNA, CYP3A4	GPR124
Stanozolol (DB06718)	Stanozolol Binding Protein (STBP)	
Fulvestrant (DB00947)	ESR1, CYP3A4	GPER, CDH2
Flucloxacillin (DB00301)	pbpA, CYP3A4	Cytokine receptors
Darifenacin (DB00496)	CHRM1 5, CYP3A4, CYP2D6	
Lubiprostone (DB01046)	CLCN2, CBR1	PTGER4, HRH2, GHSR
Ipratropium bromide (DB00332)	CHRM1 3, CYP3A4, CYP2D6	
Oxyphenonium (DB00219)	CHRM1	
Piperacillin (DB00319)	pbp3, penA, pbp2a, pbp1b	GPR124
Azidocillin (DB08795)	pbp2a, pbp1b, pbp3, pbpA, penA	
Rocuronium (DB00728)	CHRNA2, CHRM2, HTR3A	
Cefepime (DB01413)	mrdA, ftsl, pbpB,pbpA	CD248 (membrane bound signaling molecule receptor
Trospium (DB00209)	CHRM1, CYP2D6	
Estramustine (DB01196)	MAP2, MAP1A, ESR1, ESR2, CYP3A4	OPN3
Naltrexone (DB00704)	OPRD1, OPRK1, OPRM1, OPRS1, UGT1A1	
Ziprasidone (DB00246)	DRDs1 - 5, HTRs, ADRAs, CHRMs	
Ursodeoxycholic acid (DB01586)	AKR1C2, CYP2E1	Cytokine/Nuc.Harmone receptors

[176] was used to identify drugs that have similar gene expression profiles by using a large collection of transcriptional responses following compound treatments in the connectivity map project [177], [178]. Following is a comparison of some antibiotic drugs that have similarities to drugs which have GPCRs as interacting targets in their expression profiles.

Dicloxacillin is an antibiotic drug which has bacterial targets and it has similar expression profile to Metixene which is an antipsychotic drug with antimuscarinic activity by targeting muscarinic choline receptors CHRM1, CHRM2,CHRM3,CHRM4 and CHRM5. Dicloxacillin is also similar in expression profile to Velnacrine that acts as inhibitor of proteins with cholinesterase activity and have therapeutic role in Parkionson's disease (Figure 5.8).
Clindamycin is another antibiotic drug that has similar expression profile to Molindone that has DRD2 as a drug target and acts by decreasing dopamine activity. Clindamycin has also similar expression profile to Nalbuphine that exerts its action by binding to mu, kappa and delta opiod receptors in the GPCR family of drug targets (Figure 5.9).
Flucloxacillin is another antibiotic drug that has similar expression profiles to Cyclizine which posses anticholinergic, antihistaminic and central nervous system depressant properties by acting as antagonist of protein histamine H1 receptor and muscarinic choline receptors CHRM1,CHRM2 and CHRM3. It has also similarity to Prestwick-972 that has interaction as agonists of proteins with GABA receptor activity that belong to GPCRs family and are chief inhibitory neurotransmitter in the CNS (Figure 5.10).

5.4 Conclusions

The GPCR class of targets that play important role in the indication areas of central nervous system and many neurodegenerative disorders make them important therapeutic class of targets. With the availability of more structural data for GPCR targets that allow to investigate them in structure-based design methods, GPCR focused libraries will be needed to study the receptor-ligand interactions. Similarly, the availability of specialized GPCR specific ligand databases provides a means for undergoing QSAR modeling for the design of GPCR focused compound libraries. Machine learning based classification approaches are very helpful to classify GPCR binders from non-GPCR binders and can be used to design target family specific set of ligands that can be used to target a specific family of GPCRs. The classifier model built in this work can be effectively used to prefilter a database of compounds that can be specifically used in virtual screening against the GPCR targets. Using the predictive power of such machine learning based approach can provide a set of potential re-purposing candidate drugs that can be further investigated in more detail. We have identified a set of 20 known drugs that includes 6 antibiotic drugs, are proposed to be studied for activities against multiple GPCR targets. Some of the predicted drugs have similarities

to drugs that have GPCR as drug targets as discovered by their gene expression profiling. The proposed re-purposing candidates, however, need to be validated experimentally.

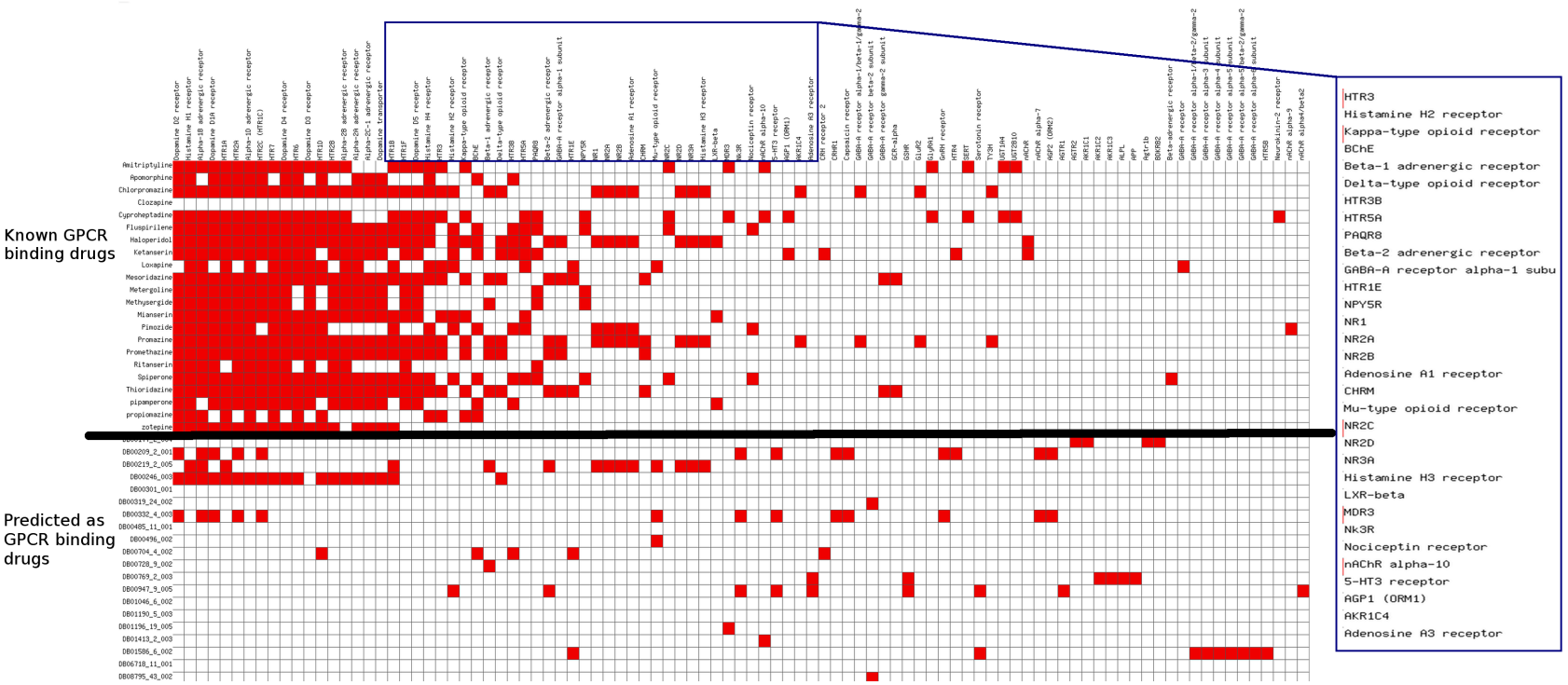


FIGURE 5.7: Heatmap of drug-target interactions predicted by MetaDrug algorithms for two groups of drugs. Drugs are given in rows and targets are given in columns. Black lines divides the drug groups into FDA approved known GPCR drugs and predicted GPCR binding drugs. Some GPCR targets are zoomed into the box.

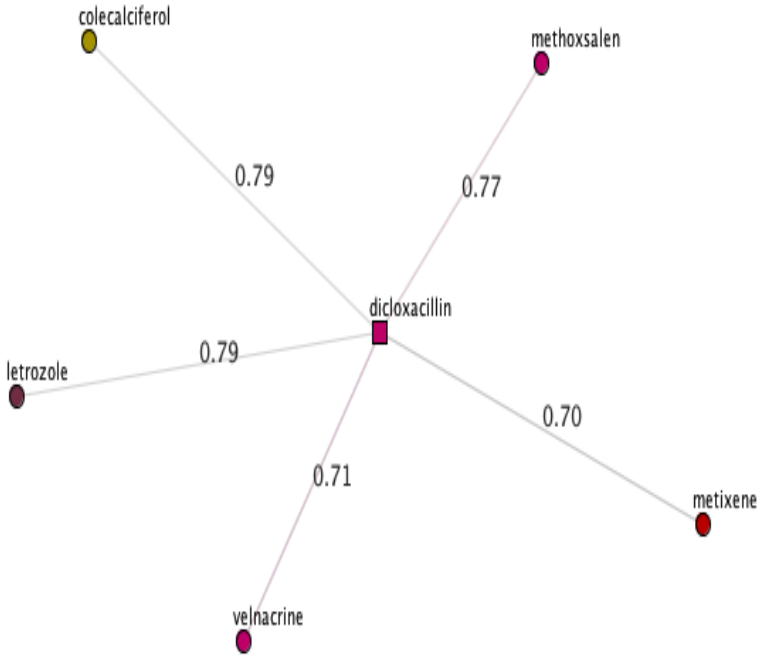


FIGURE 5.8: Drug similarity network of dicloxacillin. Drugs that have similar expression profiles are connected by edges that are labelled with similarity values.

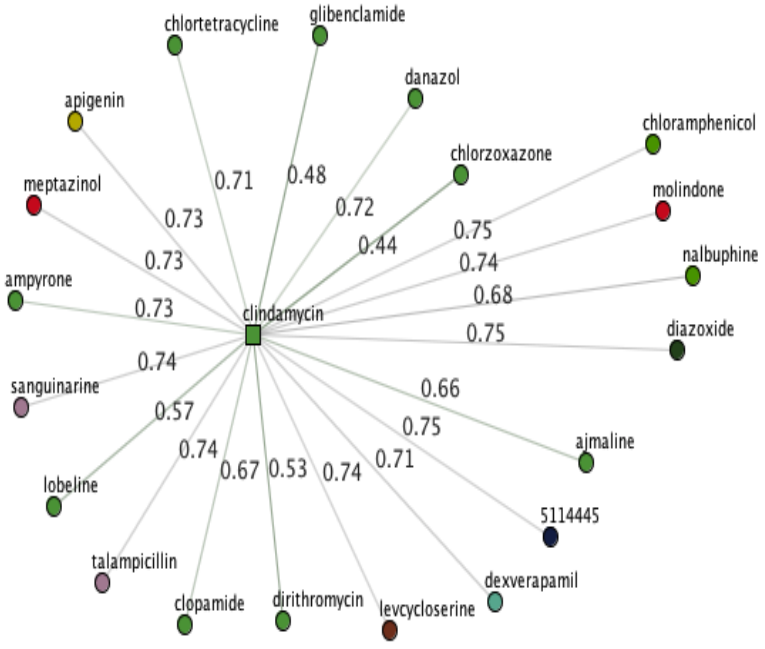


FIGURE 5.9: Drug similarity network of clindamycin. Drugs that have similar expression profiles are connected by edges that are labelled with similarity values.

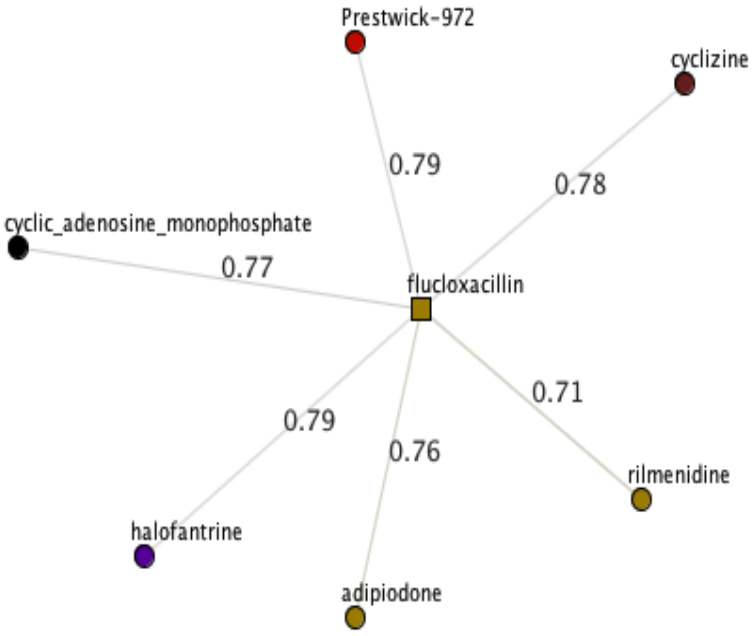


FIGURE 5.10: Drug similarity network of flucoxacillin. Drugs that have similar expression profiles are connected by edges that are labelled with similarity values.

Chapter 6

Pose Ranking and Selection to Increase Virtual Screening Accuracy and Hit Rate

6.1 Introduction

Virtual screening has emerged as an effective tool that plays important role in the discovery of novel lead compounds [179], [180]. Among the various methodologies in this area, the use of structure based design methods is increasing rapidly [181]. Having experimentally determined the protein structure of a potential drug target, virtual screening by molecular docking, in combination with many other methodologies, has become an established technique in lead identification and optimization [182]. In the molecular docking approach, the docking tools are used to predict the binding modes of the small molecule that is docked to a 3D structure of a receptor, and to score the resulting fitted complex [183]. However, the docking programs are generally more successful in predicting good binding modes as compared to correctly ranking the binders due to the limitations in their built-in scoring schemes [184], [185]. These limitations and challenges in the scoring functions and their failures in efficiently scoring the near-native binding poses have been reported often [186],[187]. Various alternate strategies and methodologies have been suggested with strong emphasis on post-docking strategies [188]. Regardless the docking programs are still frequently used as a tool for virtual compound screening and their coupling with other tools, such as molecular dynamics simulations [189] or re-scoring by using MM-PBSA/GBSA methods, have resulted in further improvements in achieving the goals [190]. Similarly, many

other approaches such as utilizing structural interaction information in the compounds screening process are gaining importance [191], [192]. Such approaches have proved very useful in the improvement of efficiency of the *in silico* predictions [188]. Among these approaches, protein-ligand interaction fingerprints have shown to enrich the screening results by re-scoring the docking poses to facilitate the filtering process [192], [193], [194], [195].

6.1.1 Protein-Ligand Interaction Fingerprints

Protein-ligand interaction fingerprints (PLIFs) are a linear arrays of bit strings that encode the presence or absence of a particular interaction between a protein and a ligand atom within a protein-ligand complex. These can be used in similarity searching to identify molecules with similar structures. Several studies have utilized the PLIFs information in improving virtual screening results by post-processing the docking poses [196], [197]. The main objective of utilizing PLIF information in screening a large database of compounds is to speed up the filtering process of biologically active compounds by reducing the burden of visually analyzing a large number of protein- ligand complexes to select the compounds with good binding modes. The filtering process becomes easier to sort out the compounds by identifying the key compound-receptor atom interactions.

6.1.2 Comparisons to Other PLIF Approaches

Our approach of utilizing PLIFs is different from that shown in recent studies [188], [193] where interactions are calculated from the protein-ligand complexes by applying geometric rules to determine interaction fingerprints and use similarity measures to find optimal interactions similar to that of a reference compound. Contrary to this, we use a simple weighted binary feature vector to identify a residue interaction that is either having a close contact interaction or hydrogen bonding interaction with the ligand. Here we present an easy and improved method of utilizing the interaction information based on knowledge of the receptor interactions with known active compounds. Weights are assigned to key interactions in order to generate a weighted-residue profile. The residue profile includes the interactions that are more frequently present or absent in the set of active compounds. The weighted-residue profile is also much simpler and very easy to calculate from the available data set of known active compounds as compared to using complex functions to calculate weighted PLIF profiles as shown in the literature [194], [195]. The weight values are simply added up to the final energy based scores already calculated by the docking programs for a predicted docking pose resulting in a

so-called Z-score. Thus, the built-in scoring function is not completely ignored, but rather the individual key interactions weight scores from the profile are used as an additional piece of information to rank the compounds.
This approach is an improvement to other approaches because we do not only consider the structural interactions but also take the significance of the interactions into account. Another improvement is that the residue profiles are created in a simpler way by the knowledge derived from the interactions of a set of known active compounds with the receptor. This approach, which is based on a set of many active molecules, helps to solve the problem of selecting a single molecule as a best reference in the studies of protein—ligand interactions. Moreover, different profiles can be used for different docking programs which can be interactively optimized and further improved. Interactive improvement of the profile is based on visualizing the interactions of known active and inactive compounds with the target receptor under study and identifying the key interactions that are shown only for the active compounds.
The main focus of this study described in this chapter is to demonstrate an easy and effective way of post-processing the virtual screening results. This chapter includes the methodology of utilizing PLIF information to achieve gain in accuracy and increase in hit selection rate. The improvements in virtual screening results by this methodology are described for a human GPCR target A2A and the method is also applied on a few more target proteins. The method is also compared to a known methodology that also utilizes PLIF information for pose ranking and selection. Finally, the general applicability of the method is discussed.

6.2 Methods and Material

6.2.1 Preparation of the Residue Profile

Given an experimentally known ligand-receptor interaction, an initial list of the residues in the active site pocket that are involved in ligand binding is prepared. Such interaction information can be easily obtained either from the available structural data of the X-ray structure (co-crystallized protein-ligand complex), LIGPLOT [198], PDBSum [199] or other tool [200] that can determine or visualize the interactions. Once the key residues that are involved in the interactions are identified, a simple residue profile is created from these interactions as shown in Table 6.1.
Residues in the profile are listed with their interaction types and their initial weight values which are set primarily to 1 for all interacting residues. A residue profile is

just a plain text file containing the residue names and the interaction types. As shown in Table 6.1, the weight values are equal in the initial profile, which are not taken into consideration at this moment. We set it to 1 because we do not know the relative importance of each residue in this list.

TABLE 6.1: A sample initial residue profile created from the interactions listed in Figure 6.1. Interaction types and their initial weights are listed in front of each residue.

Residue Number	Interaction type	Weights
LEU167	Contact	1
PHE168	H-Acceptor	1
PHE168	Contact	1
GLU169	H-Donor	1
GLU169	H-Acceptor	1
ASN253	H-Acceptor	1

6.2.2 Generation of Protein-Ligand Interaction Fingerprints

Protein-Ligand Interaction Fingerprints are generated from the interaction information of the top best docking results produced by the two docking programs FlexX [201] and AutoDock [202]. We used these tools successfully in our previous projects [40], [203]. Both, FlexX and AutoDock are amongst the most widely used docking tools [204]. FlexX is based on an incremental fragment construction and a conformation sampling algorithm, while AutoDock is based on Lamarckian genetic algorithm. FlexX can, by default, generate the interaction information of the predicted binding poses along with the binding energy scores of the docked conformations of the ligand. However, AutoDock does not produce such interaction information by default and therefore, AutoDockTools [205] scripts were modified to produce similar interaction information for the docked conformation of the ligand like those of FlexX. The interaction information is written to a plain text or comma separated file that contains the list of interacting residues and interaction types. The programs SKIFP and SKIFP2SVG, which we have written in C/C + + language, are used to generate and visualize respectively, the fingerprints from the protein-ligand interaction information from the text/CSV files produced by FlexX/AutoDockTools script.

The initial list of residues is obtained from the available interaction data of the co-crystallized ligand bound in the X- ray structure (PDB- Code: 3EML) of A2A receptor [206], a G-protein coupled receptor (GPCR) subtype which is an important therapeutic target in neurodegenerative disorders [207]. A data set of about

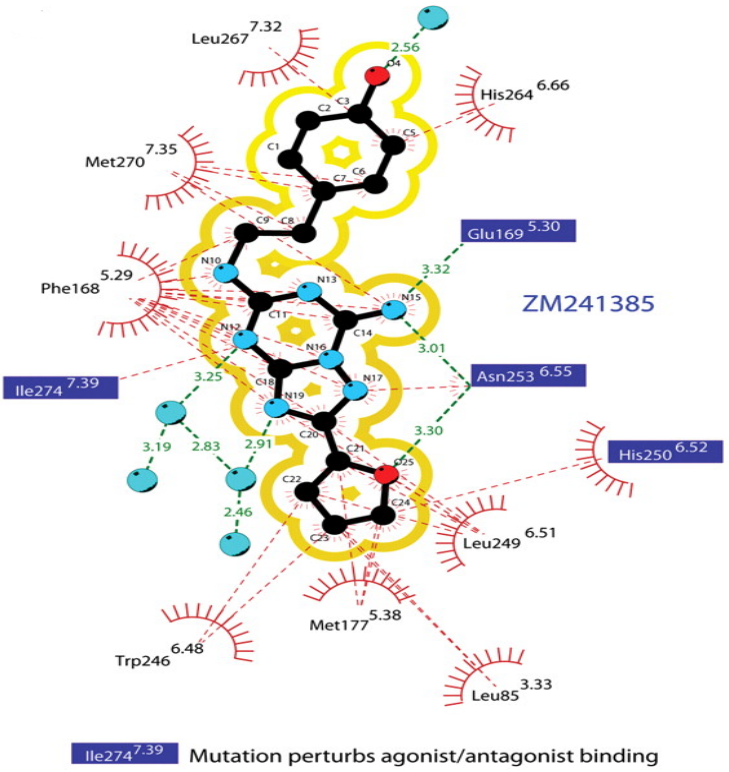


FIGURE 6.1: Schematic representation of the interactions between A2A receptor residues and ZM241385 (the cocrystallized ligand) at the ligand binding site. The dashed lines represent the close contacts and hydrogen bonding interactions of the ligand with the receptor residues and the semi-circles show the hydrophobic interactions. ZM241385 is displayed in ball-and-stick model. Residues mutations that are reported to disrupt agonist and/or antagonist binding are within blue squares [206].

160 known antagonists with experimentally determined binding affinities for A2A were collected from the literature [208], [209], [210], [211]. These known active compounds are not all similar but comprised of diverse compound classes of A2A antagonists, including adenine derivatives together with xanthine, non-xanthine and caffeine derivatives. All the compounds in the data set of known actives were docked against A2A receptor using FlexX and AutoDock and interaction information were extracted from the docking results. The fingerprints were generated and visualized by SKIFP and SKIFP2SVG programs. The compounds were sorted on the basis experimentally determined Ki values (i.e. highly active compounds at the top, low active compounds at the bottom) and not by binding energies or docking scores. This ordering facilitates building an efficient interaction profile by the identification of the key residues involved in the interaction of highly active compounds (Figure 6.2. In this figure, the top listed (80 out of 160 compounds) represent the highly actives (smaller Ki values) while the bottom listed compounds represent the less actives (larger Ki values) antagonists of A2A receptor. Each dark

band in the pattern represents the presence of the interaction of a receptor residue with a compound.

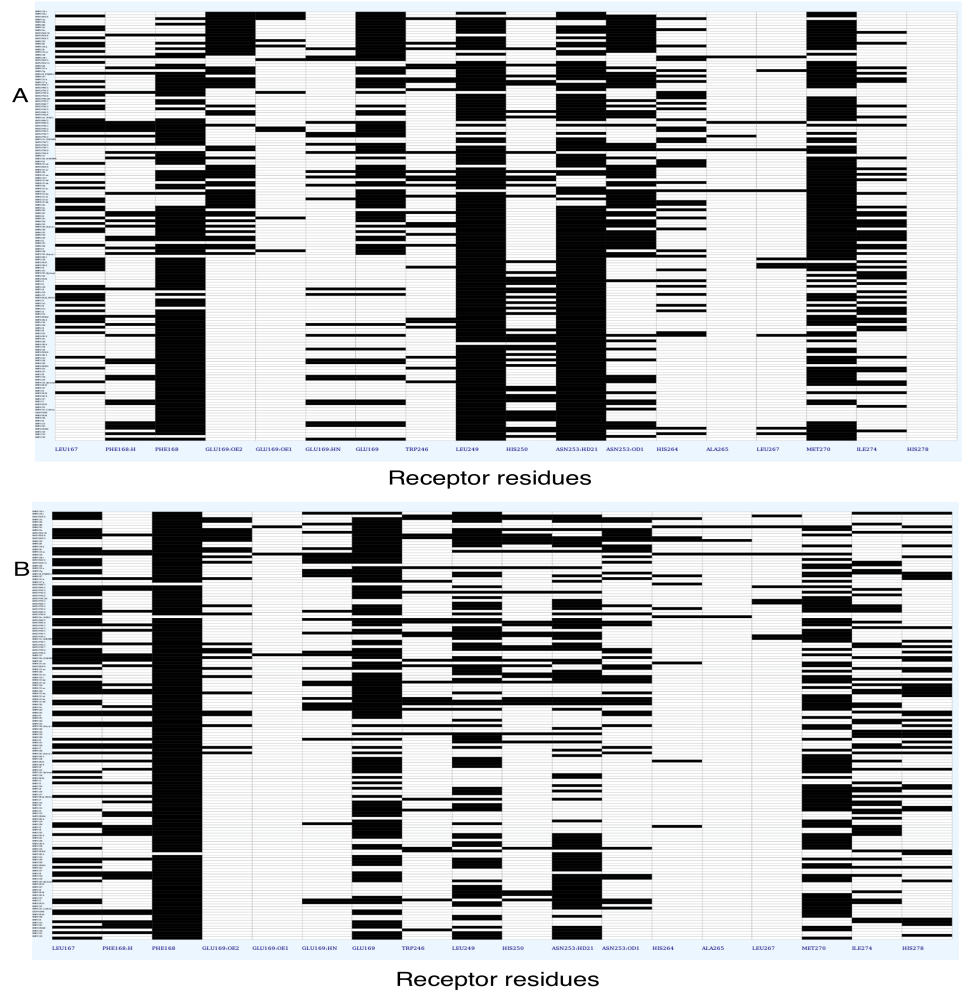


FIGURE 6.2: Fingerprint bits for the set of known active compounds by their interaction with the receptor residues given in Table 2. (A: docked with FlexX, B: docked with AutoDock). Compounds are listed on the basis of decreasing affinity, highly actives (lower Ki values) at the top and less actives (greater Ki values) at the bottom.

6.2.3 Preparation of the Weighted-Residue Profile

A weighted residue profile is prepared from the fingerprint patterns of interactions as given in Table 6.2. This method takes into account the activity data of the known active compounds. The weights in the residue profile are assigned based on the fingerprint densities of the highly active compounds. More weights should be assigned to those residues whose interaction bits are only present (darker) in the top highly active compounds subset than the less actives. A fingerprint pattern

present in all of the complete data set would mean that the interaction is present in all of the compounds and therefore is not very significant in the activity and would suggest a smaller weight value to be assigned to the respective residue in the profile (Figure 6.3).

Table 6.2: Weighted residue profiles for FlexX and AutoDock.

Residues	Interaction type	FlexX Weights	AutoDock Weights
LEU167	Contact	15	20
PHE168	H-Acceptor	1	1
PHE168	Contact	1	1
GLU169	H-Donor	30	30
GLU169	H-Acceptor	20	30
GLU169	Contact	30	5
TRP246	Contact	5	15
LEU249	Contact	5	5
HIS250	Contact	10	20
ILE252	Contact	-	5
ASN253	H-Acceptor	30	20
ASN253	H-Donor	30	30
HIS264	Contact	15	20
ALA265	Contact	15	20
LEU267	Contact	5	20
MET270	Contact	20	5
ILE274	Contact	5	5
HIS278	Contact	5	5

6.2.4 Data Sets for Method Evaluation

Two test data sets of about 1000 compounds in each set were randomly selected from the focused library and included the 80 highly active known compounds. The target-focused library was designed for the antagonist study of A2A receptor using Pharmacophore models built with MOE [92], Feature-Trees [122] similarity searching in the ZINC database [212] (that contained ≈6 million compounds) and scaffold hopping in the large combinatorial Fragment Space [213] (that contained ≈2 billion virtual compounds). These data sets are taken from the focused library instead of selecting compounds from any random "drug-like" compound library. This ensures that we avoid artificial enrichments in recovering the known actives amongst the library [214]. One data set contained compounds from the ZINC database while the other data set contained the compounds filtered from the combinatorial Fragment Space. Two different data sets are selected that are filtered

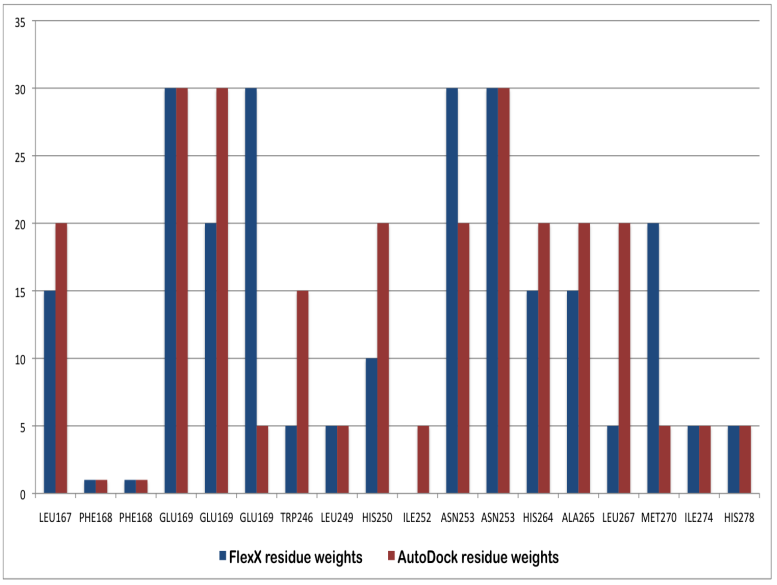


FIGURE 6.3: Weighted-residues profiles comparison used with FlexX and AutoDock dockings. Weight values are assigned to each interaction type based on the fingerprint patterns presented in Figure 6.2 for the two docking tools. With slight differences in the weight values for certain interaction types, both profiles indicate significance of key interactions in the residue profiles.

from different sources by a variety of filtering methods to evaluate the performance and stability of the method in recovering the known active compounds.

6.2.5 Virtual Screening by Molecular Docking

Docking calculations were carried out using FlexX (version 3.1.4) and AutoDock4 (version 4.1.2) for both data sets. Default parameter sets were used in FlexX dockings, while Lamarckian genetic algorithm with default parameters were used in AutoDock dockings. Rigid receptor docking calculations were used in both docking programs. Docking jobs data management, execution and post-processing of the results were performed on the compute clusters using our well-established virtual screening framework [40]. The top 10 docking predictions (i.e. poses) for each compound docked were retained for further analysis using the interaction fingerprints yielding a descriptive set of receptor residues interactions for each compound.

6.2.6 Rescoring by Z-Score and TC-MIF

Post-processing of the docking results is done by Z-scores and TC-MIF (Molecular Interaction Fingerprints with Tanimoto-Coefficient metric). For each of the top 10 docking poses, Z-scores are calculated by the SKIFP program and TC-MIF scores

are calculated by the FingerPrintLib [188] program that uses OpenEye [215] libraries. A docking pose at rank 3 or 7 out of 10 may have a docking score less than the 1st ranked pose, but may have a better binding mode if the required interactions are matched with those specified in the residue profile. A complete re-ranking is achieved for a pose by simply adding the weight values to its docking score. The Z-score, in this case, is similar to the pharmacophoric features that prioritize the docking poses based on the significant interactions specified in a pharmacophore model. However, Z-score is different than the pharmacophore model used within FlexX docking program. In FlexX, pharmacophoric constrains lead to no docking predictions if the pharmacophoric constraints need to be essentially fulfilled during the pose filtering stage.

6.3 Results and Discussion

The top 10 poses for each compound in the data sets screened against the receptor were retained for post processing. Post processing included ranking and filtering the docking poses on the basis of three criteria:

1. On the basis of highest energy/binding energy scores of FlexX and AutoDock
2. On the basis of highest Z-score by using PLIFs with weighted- residue profiles for the top 10 docking solutions, and
3. On the basis of TC-MIF values calculated for the top 10 docking solutions with using the co-crystallized ligand as a reference.

Database enrichment is calculated from the percent active molecules recovered from the percent database screened. Results quantified from both screening data sets have shown that Z-score increases the hit rates in the top 5% to 15% of the database in both docking programs for all data sets (Figure 6.4 and Figure 6.5). The hit rate increases by ∼40% to ∼3 fold in the top 5% of the database docked with FlexX and AutoDock respectively. For example, in the top 5% of the 1000 compounds database (i.e. top 50 ranked compounds), FlexX score recovered 25 active compounds, while Z-score recovered 35 active compounds (Figure 6.4, data set 2). Similarly, in top 50 ranked compounds, AutoDock score recovered 10 active compounds, while Z-score recovered 40 active compounds (Figure 6.5, data set 1). This is desirable because only a fraction of the whole database is filtered (e.g. top 1% to 10% depending on the database size) for subsequent detailed analysis. Performance measures of the three filtering criteria were also quantified by ROC (Receiver Operating Characteristics) curves as shown in Figure 6.6 and Figure 6.7.

Z-score has shown similar performance, in terms of active compounds recovery in the top ranked compounds to TC-MIF based filtering of the two data sets, and thus improves the recall ratio over the docking score based ranking in the top ranking results. Furthermore, the binding poses of the top ranked compounds are more similar to the binding pose of the co-crystallized ligand in the case of Z-score and TC-MIF than the energy based scoring. This indicates the usefulness of this approach to re-rank the compounds with good binding modes without taking the co-crystallized ligand as a reference.

The main purpose of this study is not to compare the results of different docking tools or to compare the methodologies of using protein-ligand interaction fingerprints. Instead, the current study demonstrates that PLIFs can be utilized as additional source of information to improve the docking results. The molecular interaction fingerprint method (TC-MIF) has been taken as a reference tool and used to evaluate the efficiency of the current approach for its applicability in post-processing and filtering results of large scale virtual screening. However, an improvement of this approach over TC-MIF and other automated PLIF based filtering methods is that residue profile can be interactively modified to favor certain interaction types. Filtering for the desired interaction types can be used to filter compounds on the basis of their mechanism of binding (e.g. agonists or antagonists binding mechanisms).

As a cross validation test, when equal weight values are assigned to each interaction type in the weighted-residue profile, the results are not improved in terms of hits enrichment. This indicates that certain residues interactions are more important than others and therefore a higher weight value should be assigned to them. Similarly, interchanging the profile of FlexX with AutoDock and vice versa (Table 6.2 and Figure 6.3) do not produce satisfactory results as compared to using the respective profile of the docking tool, although the results in terms of database enrichment are still better than the energy based scores. This tells about a limitation of this approach that a general residue profile should not be used for multiple docking programs.

To test the general applicability of the method, we evaluated the approach on a few more targets. For this purpose we had the inhibitor data set for MAOB from BindingDB [62] and additionally selected three more targets from a DUD [173] data sets. The three targets (ACHE, ACE and INHA) were arbitrarily selected from the DUD database. All the target structures and the compound databases (their respective ligands and decoy molecules) were energy minimized using MOE with MMFF94x fore field. The crystal ligands bound in the receptor complex were re-docked to reproduce the crystal structures binding modes. The RMSD

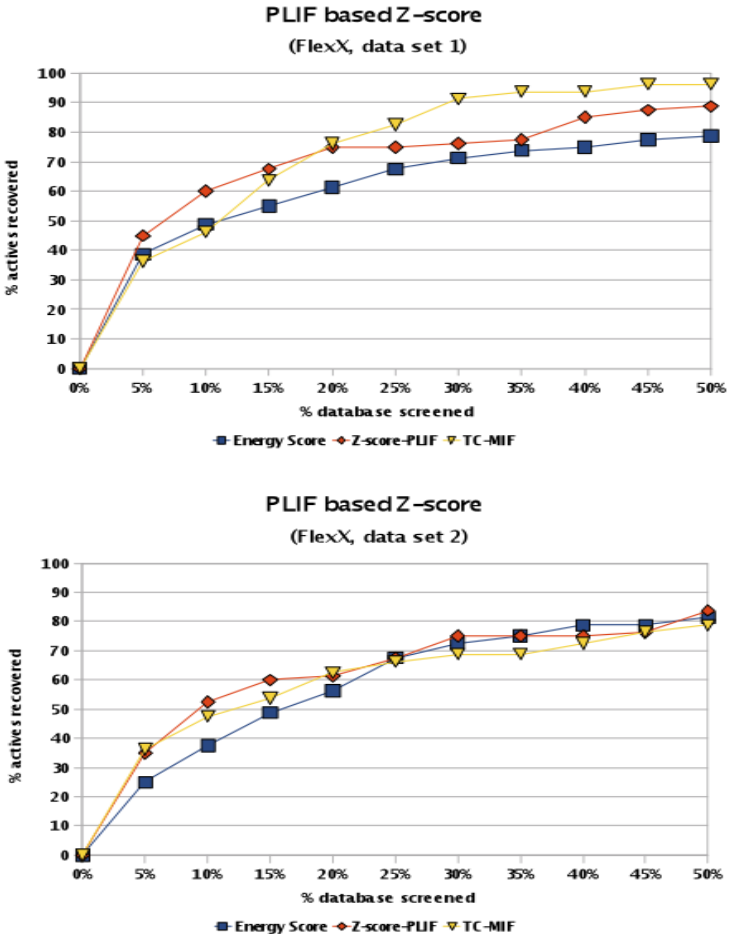


FIGURE 6.4: Z-score comparison with FlexX energy scores and TC-MIF based rankings for two data sets. Z-score is calculated from the weighted-residue profile and TC-MIF values are calculated from molecular interaction fingerprints of the docked poses in the pro- tein complex.

values for the re-docked ligands of all targets were below 2 Å. The re-docking experiment is performed to validate the docking workflow and to indicate that the target structure is prepared for docking other non-native ligands. The known active compounds of the respective target proteins were used to generate the target specific residue profiles. Afterwards, the compounds were screened against the respective targets and the results of virtual screening were ranked on the basis of docking score as well as Z-score. Improvements in terms of recovering the known ligands from the decoy database were shown in the results obtained by ranking with Z-score as presented in Figure 6.8.

In summary, the results demonstrate that PLIF based pose filtering approaches generate a new ranking as compared to the docking based score. Such approaches should be used to overcome the limitations in the default scoring functions of the docking programs. Advantages of our approach include integration of the method

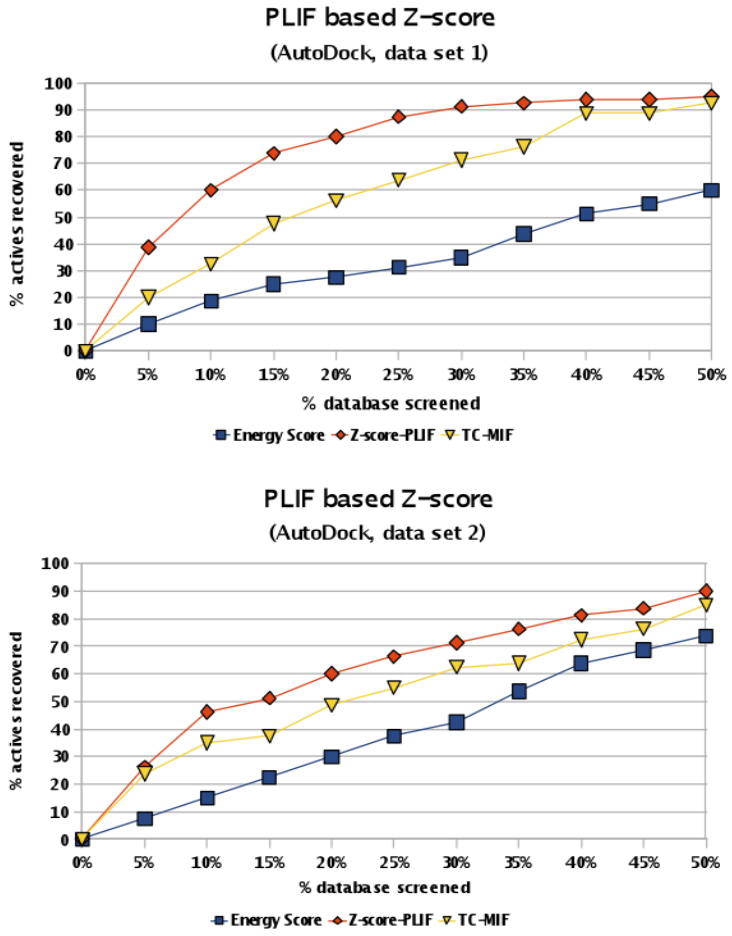


FIGURE 6.5: Z-score comparison with AutoDock binding energy scores and TC-MIF based rankings for two data sets. Z-score is cal- culated from the weighted-residue profile and TC-MIF values are calculated from molecular interaction fingerprints of the docked poses in the protein complex.

for automated post processing the virtual screening results and the potential to perform mechanism based filtering, although the mechanism based filtering is not evaluated. The docking results can be further improved if receptor flexibility is introduced during the docking calculations.

6.3.1 Binding Mode Analysis

The results ranked on the basis of three criteria (i.e. energy score, Z-score and TC-MIF score) showed a comparative performance in re-ranking the reference compound (i.e. the co-crystallized ligand) in the top ranking hits. From the rank order of the reference compound, we can suggest that the top ranking hits will have good binding modes that are similar to the reported binding mode of the reference compound. Table 6.3 shows the ranks for the compound ZM241385

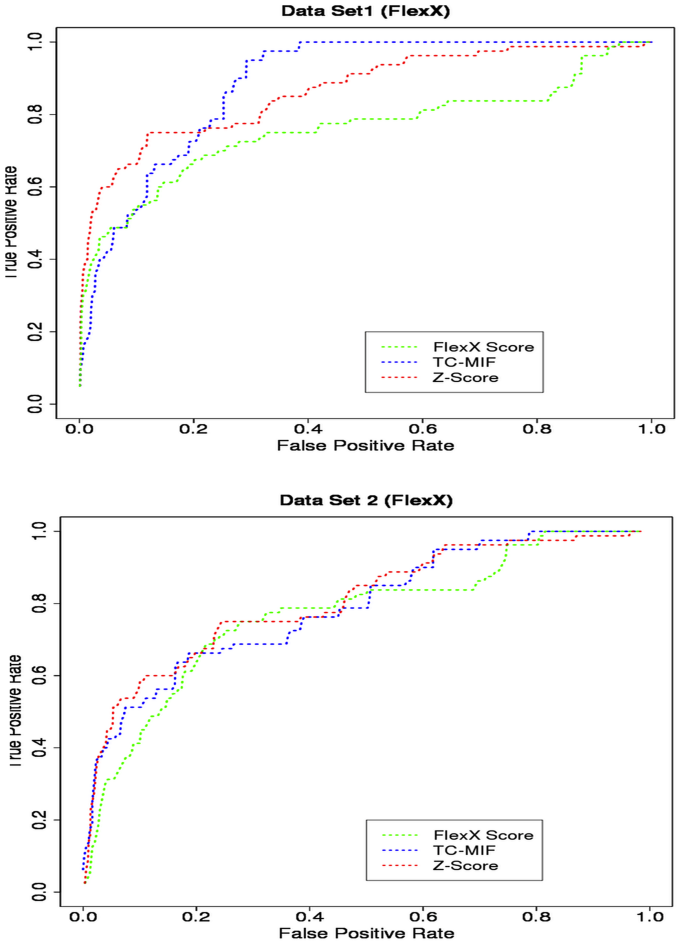


FIGURE 6.6: Performance measures of the three filtering criteria used in screening two data sets with FlexX. In both data sets, Z-score has greater recall of the active compounds than TC-MIF and Energy based rankings in the top 5 to 15% of the filtered database as also shown in the enrichment plots.

in 1000 database compounds by the three scoring criteria. ZM241385 is the co-crystallized ligand in the crystal structure of A2A receptor and is a selective A2A antagonist [216]. The rank assigned to the reference compound highlights the increase in accuracy of binding mode because it has been ranked among the top 1% of the database screened in the case of ranking by Z-score and TC-MIF in all data sets as compared to docking score based ranking. This is obvious because the PLIF based Z-score and TC-MIF score use the same interactions observed in the co-crystallized ligand as reference. However, we can only suggest a similar binding mode for the top ranking compounds but not necessarily a similar bioactivity unless it is validated experimentally.

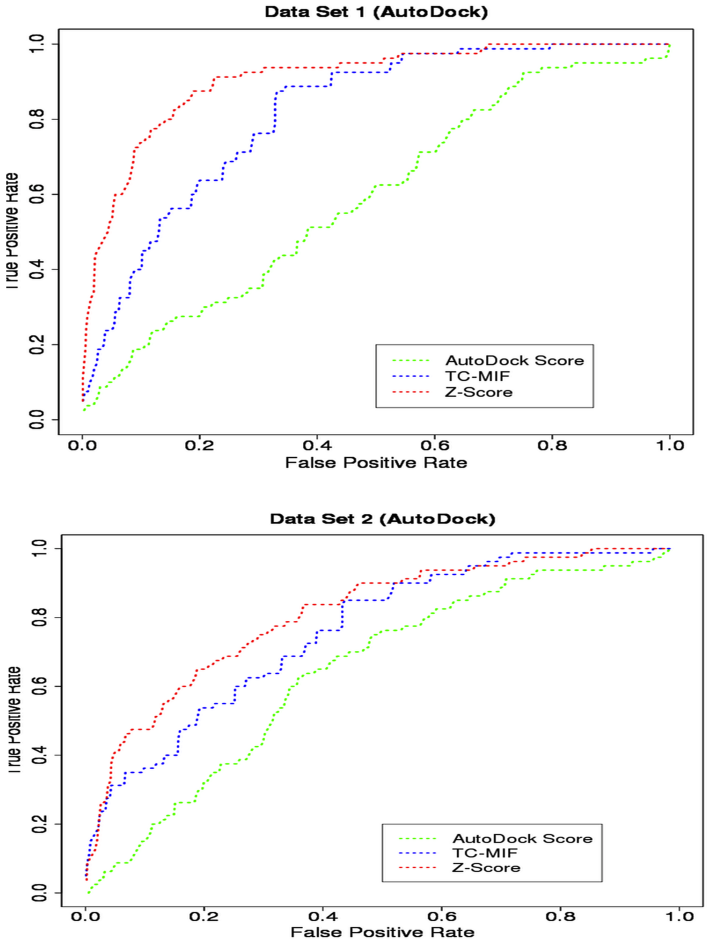


FIGURE 6.7: Performance measures of the three filtering criteria used in screening two data sets with AutoDock. In both data sets, Z-score has greater recall of the active compounds than TC-MIF and Energy based rankings in the top 5 to 15% of the filtered database as also shown in the enrichment plots.

TABLE 6.3: Reference compound ZM241384 ranked in two screening data sets by three criteria.

Criteria	FlexX Docking		AutoDock Docking	
	Data set 1	Data set 2	Data set 1	Data set 2
Energy score	15	32	208	230
Z-Score	1	3	9	13
TC-MIF	3	2	1	1

6.4 Conclusions

Docking scores are a typical way to sort out and select the best hits while analyzing virtual screening results. However, the inability of scoring functions to accurately score and rank drug candidates in a virtual screening campaign is currently the main limitation of docking tools. On the other hand, the docking programs are

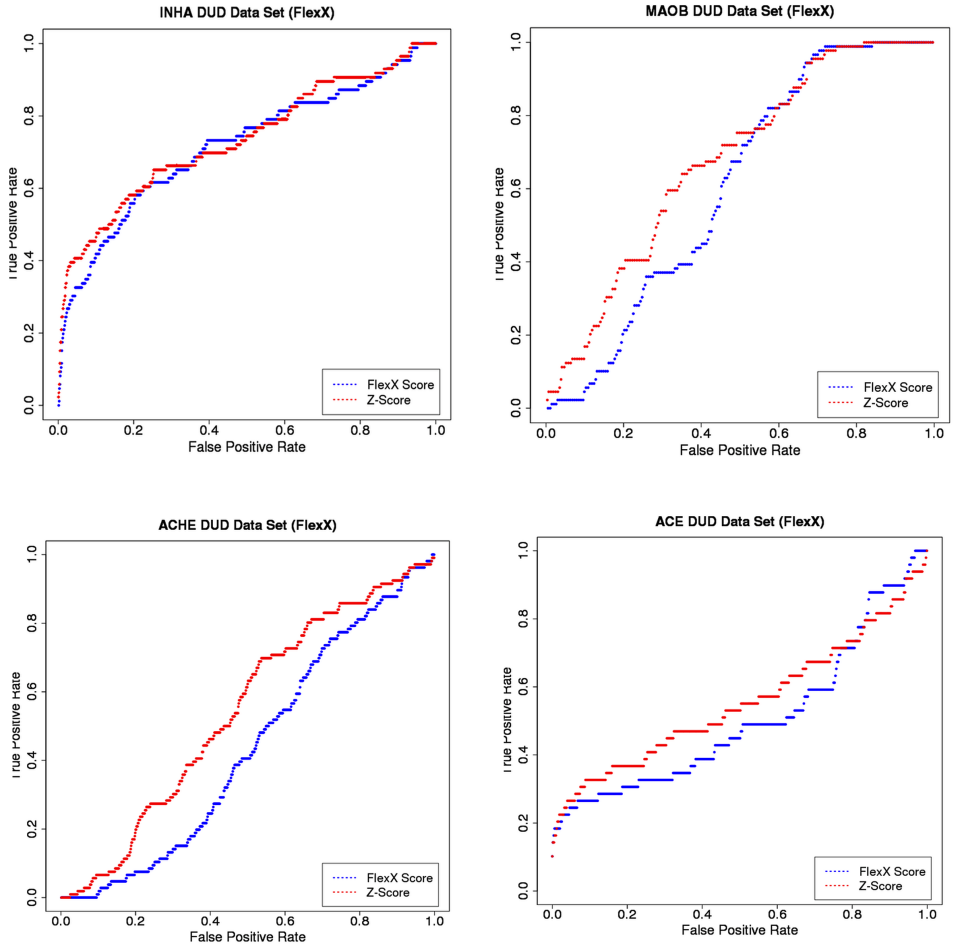


FIGURE 6.8: Improvements in recovering known ligands from the database of decoys by ranking with Z-score. Top left: INHA (Enoyl ACP Reductase), Top right: MAOB (Monoamine Oxidase-B), Bottom left: ACHE (Acetylcholine Esterase), Bottom right: ACE (Angiotensin-Converting Enzyme).

generally good in providing good binding poses. Thus, by taking the available knowledge of the protein 3D structure, its binding site and the ligand-receptor constructs, selection of the candidate compounds from the docking results require a lot of human post-processing that is time consuming. To help address this problem, we have developed a simple and semiautomated way of generating protein-ligand interaction information from the docking results. This information can be utilized efficiently in the analysis of large-scale structure-based virtual screening results to prioritize candidate compounds with good binding modes. The PLIF derived Z-score, which is based on the weighted- residue profile, is obtained for a set of important residue interactions and then added to the docking score for a predicted docking pose. In this study, Z-score has demonstrated sufficient increase in the enrichment of active compounds against the randomly selected compound data set. The data sets were obtained from the already prepared focused library

of A2A receptor. Compared to docking-score alone (i.e. energy based), the additive Z-score methodology performed very well in retrieving the active compounds against the database of decoys. In contrast to the protein-ligand interaction fingerprints that were proposed previously (e.g. TC-MIF), the methodology proposed in this study can be easily implemented in an existing virtual screening workflow, which further makes this method easily applicable by structural biologists and medicinal chemists. Though the importance of PLIF in rational drug discovery was identified long ago [192] ,[194] and several other studies published along the same lines [185], [193] the current study differs by deriving PLIFs from the generated docking results, rather than from the explicit use of protein-ligand complexes. Furthermore, fingerprint bits are residue based and not determined from the atom to atom contacts in the molecule complex. Weights in the weighted residue profiles are determined from the initial fingerprints data of known highly active compounds instead of using any mathematical function to calculate weights as presented in [195]. The weights derived in this approach are simply based on the user's preference for certain interactions. Similarly, an another advantage of favoring specific binding mechanism can help filter the results for specific binding modes. However, this also imposes a limitation on building a residue profile that can favor multiple possible binding modes. Another limitation of the approach is the requirement of bioactivity data to signify certain interactions, which is sometimes missing.

The use of PLIF information can significantly enrich the number of positive hits as compared to the number of decoy compounds in a virtual screening experiment by molecular docking. The amount of time spent and efforts involved in filtering by manual post-processing will be effectively reduced by integrating the current approach in the virtual screening pipeline. We therefore propose to integrate this approach at the crucial stage of analyzing and filtering the results of virtual screening by molecular docking.

Chapter 7

Exploring Polypharmacology through *In-Silico* Pharmacology Approaches

7.1 Introduction

Computer aided molecular models based on novel approaches towards drug discovery and development can speed up and enhance productivity while reducing costs. Such models that have predictive power can be useful for establishing efficacy and safety profiles of new drug candidates. Computer aided molecular modeling started to develop some decades ago with the enhancement and capabilities of computational methods and hardware [217]. These *in-silico* computational methods have increasingly developed over the past decades that is providing support at every stage of the drug discovery development workflow through integrating and analyzing heterogenous data in the field of computational drug discovery [218], [219]. Heterogenous data in this field are used in *in-silico* pharmacology for the creation of computational models or simulations. These simulations can be used to suggest hypotheses and make predictions for the discoveries of new therapeutics [220]. Computational methods can be applied at various drug development stages with the emphasis to reduce the time and experimental costs [221], [222],[223] that are involved at different stages of the drug development pipeline [224],[225]. Computational methods have been successfully applied in hit identification early in the drug discovery process [226], [227],[228] and have gained popularity in both academia and pharmaceutical industry, as evident in many journals that are reporting theory usage in drug development [229]. However, there is still a need to re-examine the underlying principles in *in-silico* approaches [230] with the goal of

the success rate of new drugs development [231], [232].
Virtual Screening (VS) [233], [234] is one *in-silico* approach for finding lead compounds that is complementary to high-throughput screening [235]. The major objective of VS is to accelerate the drug discovery process [236]. It is often used to identify novel chemotypes [237] and to enrich a library of compounds that have an increased chance to be active against certain molecular targets [238]. Integrated target based [239] and ligand based [240] VS have emerged resource-saving techniques that have been successfully applied in the screening of bioactive molecules [241]. VS is divided into two classes namely ligand-based and target-based VS. Ligand-based VS requires the structural information and properties of the ligands and target-based VS requires 3D structural information of the targets, and the goal is to identify bioactive ligands that will bind to the target structure [220]. Identification and validation of a successful target is one of the most important steps in new drug development [224]. To maximize the success rate of new drug development, rational drug design [242] approaches rely on the discovery of successful drug targets [243]. However, the process of new target discovery and validation is complex and bear a high degree of uncertainty [244]. Therefore, substantial improvements in the underlying target discovery approaches are needed. On the other hand, major improvements are needed in the approaches towards rational design of new drug candidates [245],[246]. A frequent argument against *in-silico* hit identification and optimization approaches is that finding hits is easier as compared with the later stage challenges faced in drug discovery [247], However, these approaches certainly allow us to cost-effectively streamline the drug discovery process [248]. Improvements and advances in both structure-based and ligand-based drug design methods [249] will have a dramatic impact on advancing new drug candidates to the clinic [250].

7.1.1 Polypharmacology

Traditional drug design strategies have predominantly been developed with the goal of identifying high affinity binders by high-throughput experimental screening of large chemical libraries, or by rational approaches based on the three-dimensional structure of the target [251]. However, in the recent years several alternative strategies, such as polypharmacology and systems biology, have emerged that will likely affect the future of small molecule drug discovery [252]. Polypharmacology has appeared in the form of drug promiscuity, where a drug hits multiple targets of the same family (e.g. kinases) [253] and multiple targets of different families [254]. This paradigm shift from the classical approaches, which are based on

target selectivity, towards polypharmacology is supported by increasing evidences [255],[256] that a drug can hit more than one target [76]. In contrast to the design of selective ligands, therapeutic polypharmacology resulting from one drug binding to multiple targets contributes to overall effectiveness of a treatment [90]. There is emerging evidence that polypharmacology resulting from compound promiscuity is key reason to the efficacy of many drugs [257], [258], [259]. A successful approach has been reported that is based on automated design of ligands for discovering their polypharmacological profiles [260]. These promiscuous drugs elicit their therapeutic effects by acting on multiple targets [261] and their promiscuous chemotypes are explored by studying them in combination with multiple targets [262],[263]. The polypharmacological relationship between ligand structures is discovered by identifying promiscuous scaffolds or "privileged" substructures present in these ligands [264]. Polypharmacology is also investigated on the target protein structures by exploring the backbone skeletal structures of protein pockets [265] to derive pocket similarities. The similarities in drugable pockets are also used to classify target proteins having similar binding sites [266] that can be used to predict the polypharmacology on the targets side. Therefore, promiscuity is not only present in the ligand structures, but it also exists in the target structures. This can help explain molecular recognition and functions of the target proteins [267] that can be exploited for polypharmacology.

There exist two definitions for polypharmacology. In the first type, there are multi-target drugs, where single drugs act against multiple targets and in the second type, polypharmacology results from multiple concurrent therapies (combination therapies) that involve the use of multiple drugs to treat complex diseases [268]. The combination therapy can not be assumed to be completely safe because of drug-drug interactions [269], [270], [271]. In this study, we will not focus on this type of polypharmacology involving multiple concurrent or combination therapies, but rather explore the promiscuity of a single drug acting against a set of multiple target proteins.

7.1.2 Combining Docking and Molecular Dynamics

With the advancements in chemoinformatics and computational techniques, docking and molecular dynamics (MD) simulation studies have become widely used in rational drug design [189], [272],[273] to identify novel ligands with detailed binding modes and interactions [274], [275]. Tools and methodologies should continue to improve so that emerging research area such as polypharmacology, can be explored [276]. Molecular dynamics is one such method that can be applied

to predict properties of small molecules, which follows molecular behavior and properties of molecules (e.g. biological or chemical entities) through time. The underlying molecular behavior requires a deeper understanding of their dynamics. Molecular mechanics (MM) and quantum mechanical (QM) calculations are often coupled with docking and MD to accurately predict binding affinities [277]. QM calculations are used for geometry optimization and MM minimizations are used to relieve the close non-bonded conflicts in the protein-ligand complex and free energy calculations.

A typical structure-based virtual screening workflow starts with obtaining a 3D structure of a protein target and a target focused compound library. Machine learning based methods are gaining popularity to filter a database of compounds [278] enriched for specific area of interest [39], [279]. Docking is generally used to obtain a favorable binding mode and to filter a large compound library by using internal scoring functions and/or subsequently enriched by using external scoring functions such as NNScore [280] and refinement with MD calculations [189]. To overcome the limitation of docking scoring functions, MM based post-processing methods serve as an automated procedure to re-score the docked protein-ligand complexes[281] and thus obtain a better ranking ligand [282]. MD simulation is extensively used to study a pre selected set of compounds for conformational flexibilities and allow to select a reasonable ligand conformation followed by the application of MM based methods (e.g. MMPBSA) for the prediction of binding affinities. Combined with docking and MD, various methods such as molecular mechanics Poisson-Boltzmann Surface Area (MM-PBSA) and Generalized Born Surface Area (MM-GBSA) are quite useful and effective in the prediction of free energy of binding [283], [284]. MM-PBSA and MM-GBSA based methods use molecular mechanics calculations and continuum solvation models to estimate binding free energies [285]. MD simulations are among important computational chemistry methods that are used for the investigation of the dynamic behavior of candidate drugs in the early stage drug discovery and also continue to develop in the future for better understanding the human biology [286]. However, MD simulations are time and resource expensive calculations, but with the availability of modern hardware such as graphical processing units (GPUs), high-throughput MD is now possible that is leading MD simulations as a standard tool in biomedical research [287].

This chapter focuses on exploring polypharmacology through *in-silico* pharmacology approaches. The approaches include the application of *in-silico* methods and tools to explore targeted polypharmacology in the area of neurodegenerative

disorders (NDD). Both ligand-based and structure-based design methods are studied. Multiple target proteins that are selected through network pharmacology, are studied and evaluated for their polypharmacology and new multi-target ligands are identified. Known drug-target interactions are studied through molecular docking and molecular dynamics simulations to reproduce experimentally determined interactions. Focused libraries of compounds are generated using chemical similarity searching and through the application of machine learning based classifier models as discussed in Chapter 4. MM-PBSA/GBSA based methods are used to evaluate binding free energies for the focused libraries against multiple targets. Further *in-silico* predictions methods based on QSAR models are performed to reproduce similar results.

7.2 Methods and Material

Know drug-target interactions are studied for exploring potential polypharmacology. A set of protein targets and drugs is selected that is used subsequently in candidate drug discovery by targeted polypharmacology. The protein targets are selected by using network based approach and used in the structure-based virtual screening. The selection of targets is critical and they should be involved in neurodegenerative disorders. The rational behind the selection of targets is based on their topology in the drug target network as described in detail in Chapter 3, and they participate in a therapeutic role in neurodegenerative disorders. There exist more than 400 FDA approved targets present in the NDD drug-target network that are therapeutically relevant in neurodegenerative disorders, as mentioned in Chapter 2. Therefore, it is necessary to define criteria that allow for prioritizing targets for our study of polypharmacology. The targets, which are selected out of those 400 available targets in the network will serve a representative set of targets for exploring polypharmacology. Table-7.1 shows a list of target proteins and their associated major disease categories that were selected for virtual screening based on network pharmacology approaches.
The rational approaches towards target selection, as described Chapter 3, consist of various criteria that are briefly described as follow:

1. The selected targets should be validated and approved targets with clear therapeutic role in neurodegenerative disease indications.

2. The selected targets are interacting with FDA approved drugs that are currently used for the treatment of neurodegenerative disorders.

TABLE 7.1: List of selected targets and their associated disease classes retrieved from the curated CTD[49] database.

Target	Full Name	Disease Categories
ADRB2	Beta-2 Adrenergic Receptor	CNS[a]/Immune, Cancer
ADORA2A	Adenosine A2A Receptor	CNS/Mental, Cardiovascular.
DRD3	Dopamine D3 Receptor	CNS/Mental
HRH1	Histamine H1 Receptor	CNS/Immune, Cardiovascular
CXCR4	Chemokine CXC4 Receptor	CNS/Cancer
PDE4A	Phosphodiesterase 4A	Mental disorders
MAOB	Mono-amine Oxidase B	Blood/Cardiovascular, Mental/CNS

[a] CNS: Central Nervous System.

3. The targets have experimental three-dimensional structural data available.

4. The targets exhibit polypharmacological profiles: they are one of multiple targets that interact with an approved drug used for the treatment of NDDs.

5. The selected targets have significant binding site similarities.

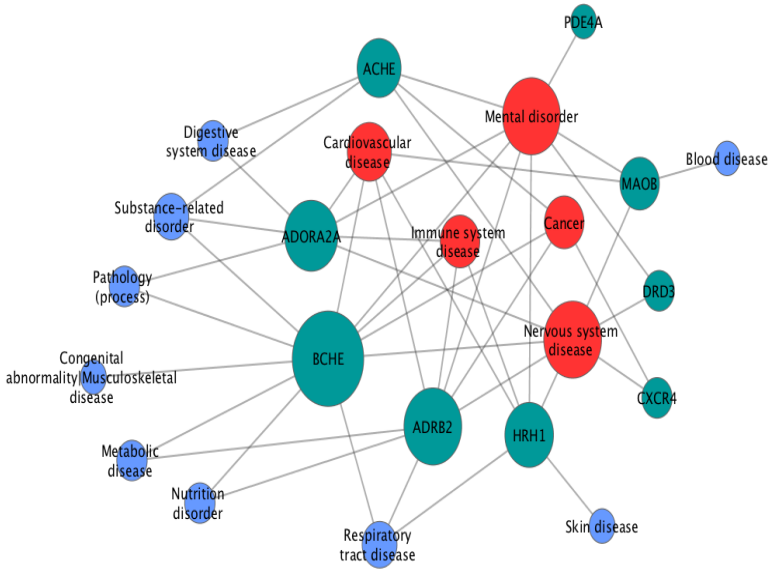


FIGURE 7.1: Disease-target network generated from curated gene-disease associations retrieved from CTD database. Target genes are shown in green color and disease categories are shown in red and blue color. Disease categories shown in red color highlight that most of the selected targets have therapeutic roles in the these disorders.

A disease-target association network is depicted in Figure-7.1, which exemplifies the importance of selected targets. The selected targets were queried in the CTD database for gene-disease associations and their curated disease categories were

retrieved. These targets have therapeutic role in various diseases including neurodegenerative and other CNS related disorders, as shown from the disease-target network built from CTD's gene-disease associations. Most of the selected targets are G-protein coupled receptors (GPCRs), since many GPCRs lack crystal structures, we selected targets that have X-ray crystal structure available to facilitate the structure-based molecular modeling studies. There is significant binding site similarity between the selected 5 GPCR targets mentioned in Table-7.1 and therefore, the selected target set represents a good example of exploring polypharmacology based on binding site similarity. Another reason for selecting this set of targets was that they are the primary and secondary targets for FDA approved drugs that are currently used in various NDDs. Additionally two enzymes (MAOB and PDE4A) were also considered to validate the polypharmacology hypothesis as these two targets have also, although very low, binding site similarity with the other selected targets. According to our polypharmacology hypothesis, a set of compounds will result in similar or lower free energy of binding by interacting with those targets that have high binding site similarities. Conversely, the compounds should have dissimilar or higher free energy of binding by interacting with those targets having low binding site similarities.

TABLE 7.2: FDA approved drugs with their indication areas from DrugBank (DB), Therapeutic Targets Database (TTD) and Comparative Toxicogenomics Database (CTD), selected for polypharmacological study against the selected targets.

Drug	**Indication Areas**
Apomorphine	Parkinsons disease (DB/TTD), Epilepsy (CTD)
Bromocriptine	Parkinsons disease (DB/TTD), Seizure (CTD)
Buprenorphine	Severe pain, opioid dependence (DB), Pain (TTD), Seizure (CTD)
Cabergoline	Parkinsons (DB), Hyperprolactinemic disorders (TTD), RLS (CTD)
Dopamine	Myocardial infarction, Trauma (DB), Seizure (CTD)
Epinephrine	Anaphylaxis, Sepsis (DB/TTD), Amnesia, Seizure (CTD)
Haloperidol	Schizophrenia (DB/TTD), Alzheimers, dementia, Epilepsy (CTD)
Lisuride	Parkinsons disease (DB/TTD), Depression, Seizure, Tremor (CTD)
Pimozide	Tourette's Disorder, Antipsychotic (DB), Schizophrenia (TTD)
Naltrexone	Heroin/alcohol dependence (DB/TTD), Multiple Sclerosis (CTD)
Risperidone	Schizophrenia, dementia (DB/TTD), Dementia (CTD)

The drug-target interaction sub-network, composed of the selected FDA approved drugs around the targets, is depicted in Figure-7.2 that shows the structural similarity of drugs together with the binding site similarities of their interacting targets. The sets of selected drugs and targets represent a good selection for the study of targeted polypharmacology since many of the drugs interact with multiple targets. Such interactions are experimentally known and direct interactions. Possible interactions may also occur based on similarity theory. Moreover, many drugs have reported experimental activity data available for the selected targets as reported in BindingDB [62] and KiDB [63]. The experimental activity data was used as a reference to validate our polypharmacolgy workflow and reproduce the same interactions by estimating the free energy of binding. The drugs mentioned in Table 7.2 and depicted in Figure-7.2 have experimentally determined Ki values for ADRB2 and DRD3 receptors. For the rest of the GPCR targets, none of the drugs have experimental activity data available. Activity values of most of the drugs were also found for some other GPCR receptors not studied here, however, those targets do not have three-dimensional structure available and hence were not considered for comparison. Therefore, MD study was performed on the selected drugs against 5 GPCR targets and comparisons with experimentally determined activity values were carried out for ADRB2 and DRD3 receptors.

7.2.1 Workflow

The workflow outlined in Figure-7.3 represents the major components and models applied in the study of targeted polypharmacology. The workflow is duplicated in two parts. The purpose of the first part is to reproduce the known drug-target interactions and explore polypharmacology on a set of multiple targets. The purpose of the second part is to discover novel candidate compounds that exhibit polypharmacological profiles. The second part extends the same workflow by including focused compound libraries filtered by the classifier model (Chapter 4) and fingerprint-based similarity comparisons. The set of selected approved drugs are first screened by docking against the targets, then their best binding poses are filtered. Afterwards, MD simulations are performed for all drug-target complexes. Using MM-PBSA/GBSA based methods, the free energy of binding are estimated for the complexes.

The aggregated knowledge and data that is coming from various knowledge base resources and stored in the HBP forms a basis for designing data-driven models. The design, constructions and evaluation of computable models are used for

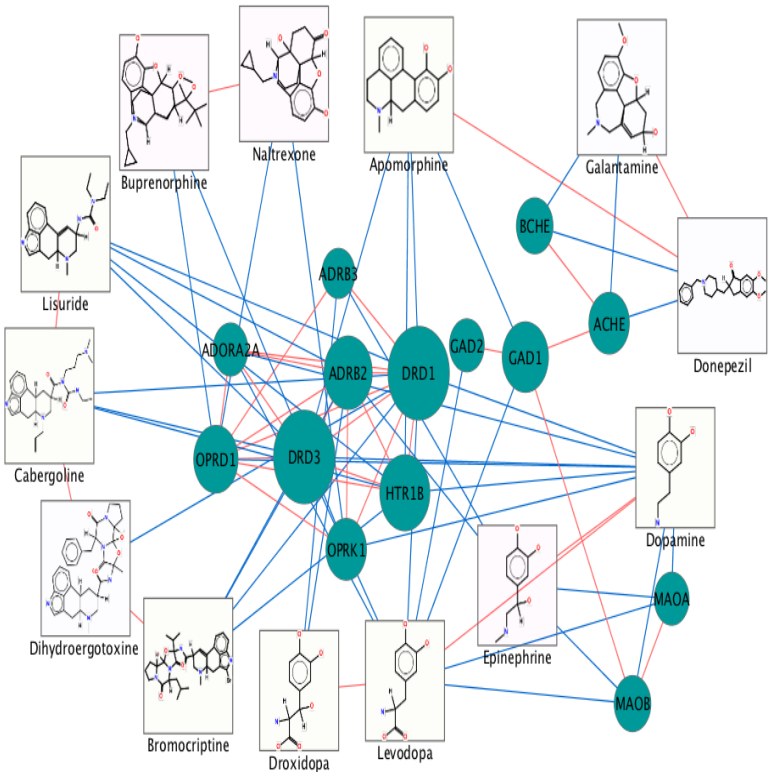


FIGURE 7.2: A set of FDA approved drugs interacting with selected targets. Red lines indicate drug-drug and protein-protein binding site similarities and blue lines indicate drug-target interactions. 2D structural diagrams are shown for the selected drugs.

studying targeted polypharmacology in brain disorders in general and in neurodegenerative disorders in particular. The workflow outlines three major modeling approaches originating from the HBP after the initial data collection step is performed. The first approach is the design of ligand-based models combining QSAR and machine-learning methods to characterize drugs and ligands for their potential evaluation in NDD disorders. With the help of this approach, existing drugs are evaluated for cross-indications (drug-repositioning), and for designing disease specific focused libraries (e.g an NDD focused library). The second approach uses network-based pharmacology to identify and select drug-target combinations for prioritizing targets to be evaluated in the study of targeted polypharmacology. The third modeling approach is exploring and validating (both *in-silico* and *in-vitro*) polypharmacology by combination of structure-based design tools and methodologies. These methodologies include combining docking and molecular dynamics into a single workflow, in which a series of structure based virtual screening tools are applied. Molecular dockings are performed on an array of selected targets, in which poses of a ligand are subsequently ranked and filtered for the respective targets. Molecular dynamics simulations are performed for the top-ranked

poses. Finally, a molecular mechanics based re-scoring is performed to obtain a free-energy ranking that incorporates the complex's dynamic motion.

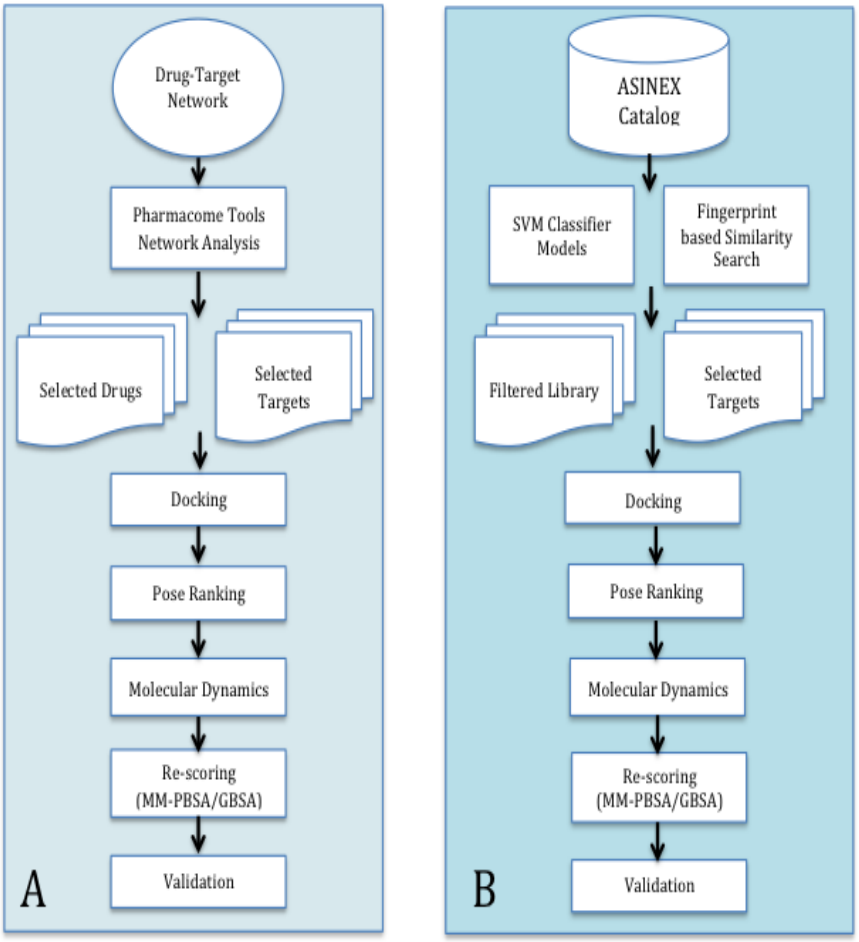


FIGURE 7.3: Workflow components for the study of targeted polypharmacology. A) Workflow adapted for drug-targets selection, parameterization of individual workflow components and validation by reproducing known drug-target interactions from drug-target network. B) Extended workflow with same workflow components but using focused compound libraries filtered through SVM based classifier model and similarity searching.

In order to perform the virtual screening experiments, it is necessary to prepare the molecular structures of both drugs and targets accordingly for use with the modeling tool. The preparation steps include generation of 3D conformations for the drug molecules and proper protonation and energy minimization of the target protein structures. The compounds and protein structures preparatory steps were performed with MOE 2011.10 [92], docking was performed by LeadIT 2.1.0 [288], pose ranking and selection was carried out using NNScore v2 [280] and molecular dynamics simulations were performed by using Amber12 [289], [290] and the inputs for MD were prepared with AmberTools-12 [289]. Binding free

energies for all protein-ligand complexes were calculated using MM-PBSA/GBSA [291] methods. *In-silico* validations for the newly predicted candidate hits were performed by comparative modeling algorithms available in MetaDrug database [175]. The major steps of the workflow are described as follow:

1. Target selection and prioritization as described in section "Methods and Material". For both workflow steps, the same set of selected targets were first used in reproducing known drug-target interactions and then for screening focused libraries for the exploration of polypharmacology.

2. For the purpose of validating the polypharmacology workflow and reproducing the experimentally determined drug-target interactions, a set of approved drugs around the selected targets (as described in section "Methods and Material") was also selected. These drugs are used for the treatment of various nervous system disorders particularly the major NDD disorders such as Parkinsons, Alzheimers and Epilepsy. The set of selected drugs with their indication areas are listed in Table 7.2 with respective references to different database resources.

3. Drugs, compound library and target structures were prepared by generating their 3D conformations, protonation at normal physiological pH, ionization and energy minimization steps. The 3D conformations were generated using MOE. Protonation, generation of tautomers, ionization states were generated by using the Protonate3D algorithm implemented in MOE. The resulting protonation states were energy minimized with MMFF94 forcefield using MOE. The target protein structures were prepared by applying protonation with Protonate3D implemented in MOE followed by energy minimization using Amber99 force field in the presence of co-crystallized ligand(s) and solvents.

4. Focused libraries of compounds are generated using machine-learning based classifier model and similarity searching. The compound libraries are generated from a commercial catalog available from ASINEX (http://www.asinex.com). Detailed description about the compound library generation is provided in the coming section "Focused Libraries Enrichment".

5. Docking of selected drugs and filtered compound library was performed against all selected targets by using LeadIT docking software. For each drug/compound, top 10 docked conformations were retained for subsequent filtering (i.e. by external scoring functions as described in the next step)

of most favorable conformation. Further details are provided in the section "Docking and MD of Focused Libraries".

6. Pose ranking and selection was performed by using NNScore algorithm, which facilitates the selection of favorable binding pose among the predicted docked poses against each target protein. The first ranking pose was selected for each compound that has the highest predicted activity against the respective protein target. Further details are provided in the section "Pose Ranking and Selection".

7. Protein-ligand complexes are prepared by using tools and utilities available in AmberTools to apply respective force fields and parameters. Detail descriptions are provided in the section "Input Preparation".

8. Molecular dynamics simulations were performed over the selected protein-ligand complexes in a series of multiple steps. Details are given in the section "Production MD Run".

9. Binding free energies were calculated by using MM-PBSA/GBSA based methods that were applied on the MD trajectories obtained from the production MD runs. The trajectories of the protein-ligand complexes were analyzed at the interval of every 10ps time as it was observed with autocorrelation function that statistically different energy values could be sampled at this time interval. The binding energies are not based on the entropy of the protein-ligand complex, but are estimated from the interaction energy and solvation energy for the complex, receptor and ligand.

10. *In-silico* validations of the known drug-target interactions observed with docking-MD were carried out by comparison to experimentally determined activity values derived from bioactivity databases. Likewise, *in-silico*validations for the interactions of novel candidate drugs were performed with QSAR modeling algorithms available in MetaDrug.

In the coming sections, a series of preparatory input setup is described for the drugs and protein target complexes.

7.2.2 Input Preparation

Before performing simulations, the protein-ligand complexes were first parameterized and equilibrated. For the ligands, general amber force field (GAFF) [292] was

used and for the protein structures, FF99SB [293] was used. No missing parameters for the ligand atoms were reported by AmberTools. The protein structure optimized with MOE was used as input for the receptor atoms having no co-crystallized ligands. The protein and ligand atoms were combined by using Leap module of AmberTools. The complexes were solvated by using explicit solvent TIP3P water model [294] in an octahedron box with at least 8 Ådistance around the complex. The total charge of the complex was neutralized by adding counter ions (i.e. Na+ or Cl-). Disulfide bonds were explicitly specified if they occur between some residues of the protein. This is important because without specifying these bonds, the protein conformation can drastically change during molecular dynamics simulation.

7.2.3 MD Equilibration

Equilibration steps include heating, pressure and density equilibration of the protein-ligand system. The complexes were minimized for 5000 steps of minimization. The minimized complexes were subjected to heating from initial temperature of 0K to 300K for 50ps with NMR restraints, followed by density equilibration at constant pressure for 50ps. The energy-minimized complex was subjected to gradual heating via the NMR option to avoid heat shocking the complex. Very weak restraints (i.e applying a weak force of 2.0 k cal/mol–$\mathring{A}^2$) were used to hold the heavy atoms of protein residues fixed during the temperature, density and pressure equilibration steps. However, no restraints were used during the final equilibration step that was carried out at constant pressure for 5ns.

7.2.4 Production MD Run

The equilibrated complexes (i.e having less fluctuations in RMSD of the backbone atoms) were simulated for further 5ns production MD runs at 300K and 1 atm anisotropic pressure scaling in a NPT ensemble. A time step of 2fs was used for the simulations. The long range electrostatic interactions were treated by the Particle Mesh Ewald (PME) method by using a nonbonded cutoff of 8 Å. The SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms. A periodic boundary condition was applied and snapshots were recorded at every 2ps time interval. This resulted in 2500 snapshots for 5ns simulation run of each protein-ligand complex. These snapshots were sampled later in performing MM-PBSA/GBSA based calculation of free energy of binding. Protein conformational changes were analyzed by plotting variations in protein backbone atoms and ligand atoms. Residue-wise fluctuations were also observed for the drug-protein

complexes in order to observe the conformational changes in the protein residues upon binding a ligand.

7.2.4.1 Free Energy of Binding

Binding free energies were estimated for all the selected drug-target complexes. Using MM-PBSA/GBSA based methods to calculate binding free energies, representative snapshots at every 10 ps were extracted from the MD simulation trajectories. The estimated free energy values were compared against experimentally determined activity values for two targets DRD3 and ADRB2. The experimental Ki values were converted to ΔG values by using the following equation:

$$\Delta G = -RT\ lnKi \tag{7.1}$$

where T is temperature in Kelvin, R is gas constant and Ki is the binding activity in molar concentration. The Ki values for the selected drugs were converted from nM to molar and then converted to ΔG with the temperature of 295 K.

7.2.5 Focused Libraries Enrichement

The NDD focused compound libraries were generated from a database that includes more than 400,000 compounds available in a commercial catalog ASINEX [295]. The generated libraries were screened against multiple target proteins by molecular docking and MD. Filtering of compound database was done by two methods as described below.

7.2.5.1 SVM Classifier based Filtered Compound Library

A neurodegenerative diseases (NDD) focused library was enriched by using machine-learning classifier model [39] built for discriminating NDD drugs from non-NDD drugs as described in more detail in chapter-3. A set of NDD relevant molecular descriptors were first calculated for all the compounds in the database and the classifier was applied to predict a set of compounds that could be relevant and useful in NDD. By using the classifier model (Chapter 4), the database was filtered and top 500 compounds that were predicted as NDD ligands with highest probability according to the classifier model, were selected.

7.2.5.2 Fingerprint Similarity based Filtered Compound Library

A fingerprint-based similarity approach was used to filter the database by using extended connectivity fingerprint (ECFP) method [296]. For each compound in

the database, ECFP fingerprints were calculated using MOE. ECFP based fingerprint similarity was used because it has been shown to perform better than other fingerprint based similarity methods [297]. Similarly, the same fingerprints were also calculated for the known NDD drugs taken as reference compounds. The database compounds were ranked according to 5 nearest neighbors (5NN) search strategy, in other words, the average similarities of the database compounds to any 5 of known NDD drugs and 500 top ranking compounds were selected

7.2.6 Pose Ranking and Selection

As the first ranked pose obtained by docking score is not always the best binding pose due to the limitation in scoring functions of the docking programs [184], [185]. To overcome this problem, the selection of the most favorable docked pose was performed by using an external scoring function NNScore [280]. A comparison of the docking score based pose ranking and external scoring functions was carried out to evaluate the performance of external scoring function. A set of known active ligands and decoy compounds sets were docked against four targets (ADORA2A, MAOB, DRD3 and HRH1) and pose filtering was performed on the basis of different scoring functions. A set of known active ligands for ADORA2A and MAOB were collected from BindingDB [62] and decoy sets were generated by DecoyFinger [298] and the sets of known active ligands and decoys for DRD3 and HRH1 were collected from GLL/GDD database [172]. The mixed ligand-decoy sets were docked against the respective targets and pose rankings were performed by the docking score, NNScore and Protein-Ligand Interaction Fingerprint (PLIF) based ZScore [41] as described in detail in Chapter 6.

7.2.7 Docking and MD of Focused Libraries

The fingerprint similarity and SVM classifier based filtering resulted in a small library of compounds (total 1000) that were used in screening by molecular docking and MD simulation. The compounds in the library were protonated by using Protonate3D algorithm in MOE, resulting in a set of 1632 molecules. The selected compounds were docked against 7 target structures (5 GPCRs + 2 enzymes) and for each compound 10 docked conformations were retained. Default docking parameters were used. Semi-flexible docking calculations were performed that consisted of rigid receptor and flexible ligands. The docked conformations for each compound were ranked for each protein target and the best pose was selected by using NNScore algorithm. Finally, compound ranks for all protein targets were

averaged. As a result, 50 high average ranking compounds were selected with favorable binding modes against multiple targets. MD simulations were performed for the selected 50 compounds against multiple targets including 5 GPCRs and 2 enzymes. In a semi-automated way, 350 cross-docked ligand-protein complexes were subjected to simulations study by MD. Each complex had a 5ns equilibration simulation performed, followed by 5ns production simulation. For each complex, 2500 snapshots collected every 2 ps were recorded in each MD trajectory for the respective protein-ligand complexes. 250 snapshots (i.e. every 10th snapshot out of the total 2500 snapshots) were subjected to evaluation for free energy of binding with MM-PBSA/GBSA.

7.3 Results and Discussions

7.3.1 Docking Workflow Validation

Docking was used to generate multiple binding poses for the ligands inside the protein binding pockets. The re-docking experiments showed that the protein structures are suitable for docking and are well prepared for use in screening a compound library. With the re-docking experiments of the selected targets, it was ensured that the protein structures' binding sites are well defined and the docking parameters governing interactions are sufficiently reproducing the experimental binding modes. Table-7.3 shows a list of selected PDB structures, their native co-crystalized ligands and the re-docking results obtained by docking with LeadIT.

TABLE 7.3: Re-docking results of co-crystallized ligand docked to the respective protein targets. The RMSD (Å) values are given for the first docking pose that is ranked on the basis of docking score.

Target	**PDB ID**	**Ligand ID**	**Docking Score (kJ/mol)**	**RMSD (Å)**
ADRB2	2RH1 [299]	CAU	-35.16	0.68
ADORA2A	3EML [206]	ZMA	-24.47	0.88
DRD3	3PBL [300]	ETQ	-18.97	0.80
HRH1	3RZE [301]	5EH	-33.72	0.73
CXCR4	3ODU [302]	ITD	-14.67	4.17
MAOB	1S3E [303]	RHP	-20.21	1.02
PDE4A	2QYK [304]	NPV	-23.54	1.13

Default docking parameters were used for reproducing the crystal structures' bound modes for all protein-ligand docking, however, with the exception of one docking parameter, the "maximum overlap volume" was increased from 2.9 Å(default)

to 5 Å to relax the protein-ligand clash penalties. The re-docking results show that the RMSD of the docked pose with the co-crystallized ligand is below 1 Å or very close to 1 Å for all the targets except for CXCR4 for which the RMSD is above 4 Å. The reason for this is due to the large size of the crystal ligand that contains multiple rings. This docked ligand adapts a conformation that is slightly shifted from the original conformation of the co-crystallized ligand, but the rings positions still overlap to produce nearly similar interactions as shown in Figure 7.4.

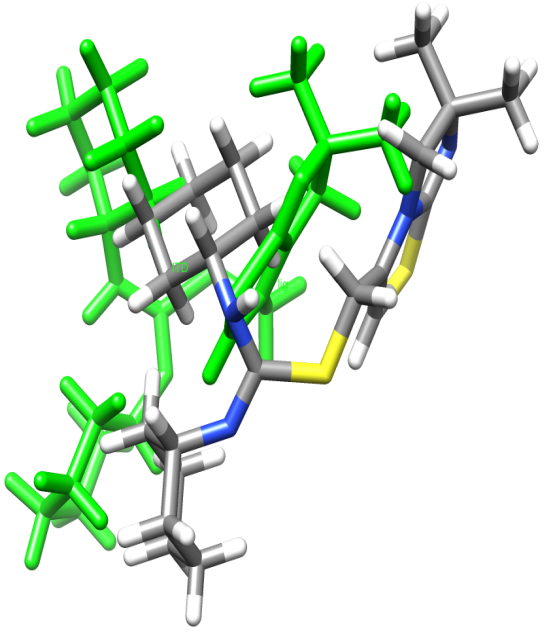


FIGURE 7.4: The re-docked conformation of the co-crystallized ligand of CXCR4. The crystal ligand is shown in green and the re-docked conformation is shown in CPK coloring.

7.3.2 Docking of Known Drugs

Docking a compound to a protein structure produces different binding modes (i.e. poses) for a single compound inside a binding pocket. Each pose is assigned a docking score according to the underlying scoring function algorithm. Due to the limitations in the scoring functions [184],[185], the best binding mode is not always the one with the lowest docking score. Therefore, an alternative method and scoring function, NNScore, was used to refine the scoring. Various binding modes of the selected drugs (Table 7.2) were obtained after docking them against the selected targets. For each drug docked to a target protein, 10 conformations

filtered by docking score, were retained for visual analysis and post-processing to select the best binding pose. Most favorable binding modes were selected that were close to the native binding modes of the respective co-crystallized ligand present in the X-ray structures of the selected targets. Visual analysis and manual selection of binding poses was performed for each drug docked against all targets and then by using pose-ranking and selection algorithm (NNScore). The binding poses selected by NNScore were similar to the manually selected binding poses in 8 out of 10 drugs that were docked against the selected targets. These observations provide a valid reason for using NNScore to filter the docking poses. Therefore, the docking scores were not considered in the binding pose ranking, using instead the NNScore algorithm to select the top ranking binding pose for each docked drug structure. The docking scores for the drugs are not given here because the scores are not used in ranking and selection of the favorable binding pose.

7.3.3 MD Simulations of Known Drugs

MD simulations were performed to evaluate free energy of binding for the selected drug-target complexes. Additionally, through MD simulations, the interactions of drug molecules with the target proteins were evaluated. MD simulations were also carried out to incorporate dynamics and relaxation, and to investigate the conformational changes on the target structures upon drug binding. The binding energy estimated for the drugs were compared against the experimental activity data for the drug-target pairs. However, not all drugs have experimentally determined affinity values against all the selected targets. Therefore, the binding energy values for only two targets were used in the comparisons for which experimental affinity values for most drugs have been reported in activity databases such as ChEMBLdb, BindingDB and KiDB.

The drugs mentioned in Table 7.2 and depicted in Figure-7.2 have experimental activity values determined for ADRB2 and DRD3 receptors. For the rest of the GPCR targets, none of the drugs have experimental activity data available for these GPCR targets. Activity values of most of the drugs were also found for some other GPCR receptors; however, those targets do not have three-dimensional structure available so far and hence were not considered for comparison. Therefore, MD study was performed on the selected drugs against 5 GPCR targets and comparisons with experimentally determined activity values for ADRB2 and DRD3 receptors.

7.3.3.1 Equilibration

Figure 7.5 shows that the density has been equilibrated in the equilibration step of 50ps density equilibration with constant volume and the density remains the same during full equilibration with constant pressure for 5ns.

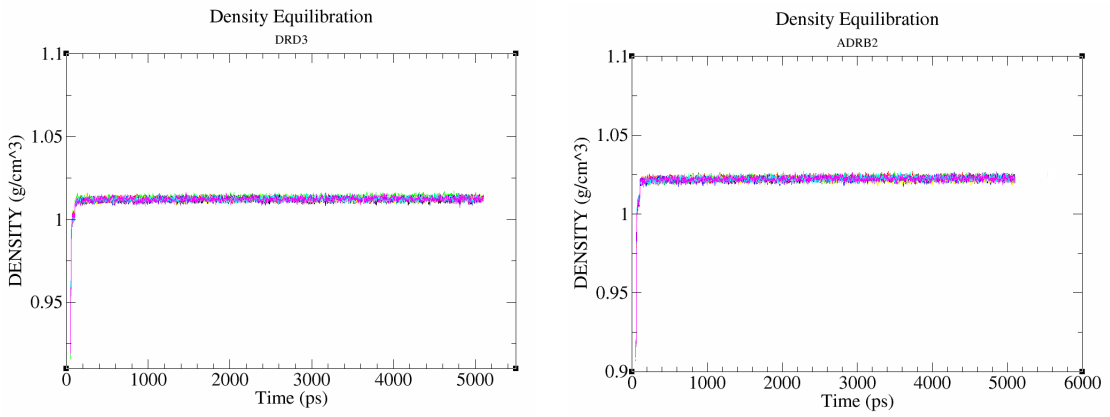


FIGURE 7.5: Plots showing density equilibration for the selected drugs (lines shown in different colors) complexed with ADRB2 and DRD3.

The protein backbone atoms changes in terms of RMSD deviations in reference to the initial docked protein-ligand complexes show that the complexes start to equilibrate between 2 and 4ns time of simulation. The equilibration results for DRD3 and ADRB2 targets are shown in Figure 7.6 and Figure 7.7 respectively. The RMSD values are reported for the protein backbone atoms only. These results indicate that a longer equilibration is necessary to obtain a stable protein-ligand complex for performing production MD runs.

7.3.3.2 Production MD Run

Production runs of 5ns period were performed after the drug-target complexes were equilibrated. As shown in Figure-7.8 and Figure-7.9, the protein-ligand complexes do not change very much during the simulation run and remain stable with slight changes in the backbone. These changes that occur in the backbone atoms are caused by the residues in the loop regions of the proteins. Residue-wise fluctuations (Figure 7.10) also indicate the conformational changes occur mostly in the intra- and extra-cellular loop regions of the protein. It can be noticed that the loop conformations changes occurring almost in a similar way for most drug-protein complexes. As shown in Figure 7.10, the residues with higher fluctuations belong to the loop structures that undergo dynamic conformational changes as demonstrated in the case of Pimozide- and Bromocriptine-ADRB2 complexes shown in

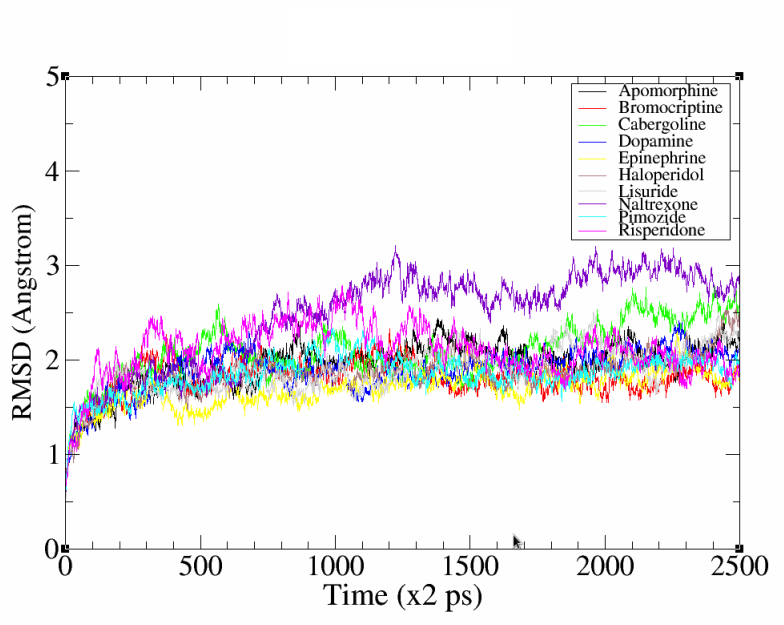


FIGURE 7.6: Plots showing backbone RMSD values for DRD3 complexed with selected drugs (lines shown in different colors) during 5ns MD equilibration period.

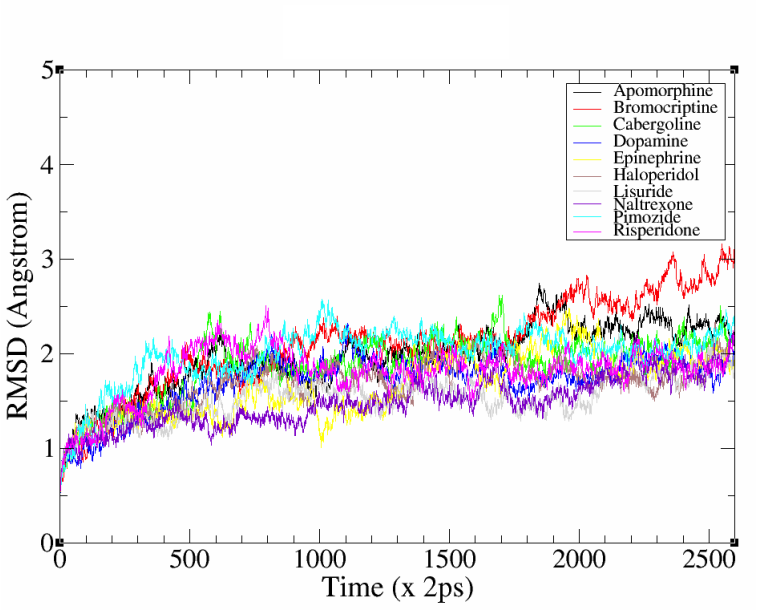


FIGURE 7.7: Plots showing backbone RMSD values for ADRB2 complexed with selected drugs (lines shown in different colors) during 5ns MD equilibration period.

Figure 7.12. It can also be observed that the drug molecules do not show conformational changes as compared to the protein atoms. The ligand atom RMSD plots are shown in Figure 7.11. This also indicates that most of the drugs have favorable binding modes for these receptors and they do not exhibit big conformational changes during the simulation. However, interatomic interactions could be

investigated to determine what particular interactions are favored by these drugs.

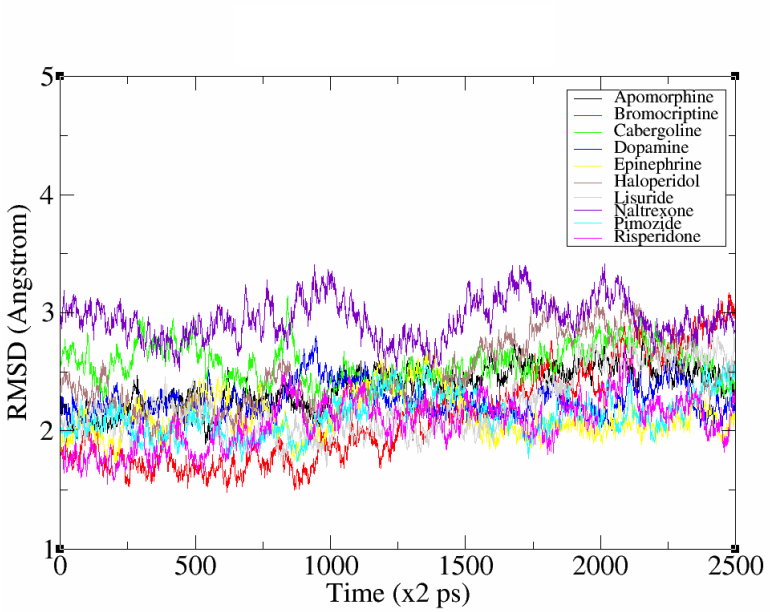


FIGURE 7.8: Plots showing backbone RMSD values for DRD3 complexed with selected drugs (lines shown in different colors) during 5ns MD simulation.

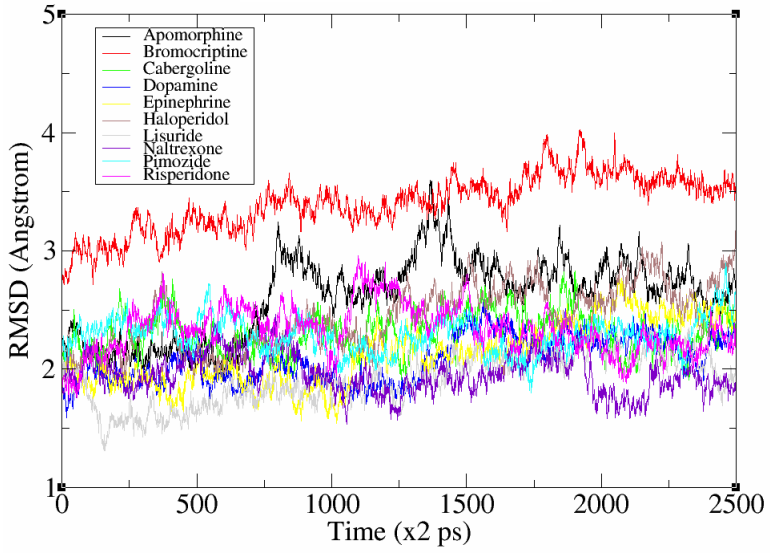


FIGURE 7.9: Plots showing backbone RMSD values for ADRB2 complexed with selected drugs (lines shown in different colors) during 5ns MD simulation.

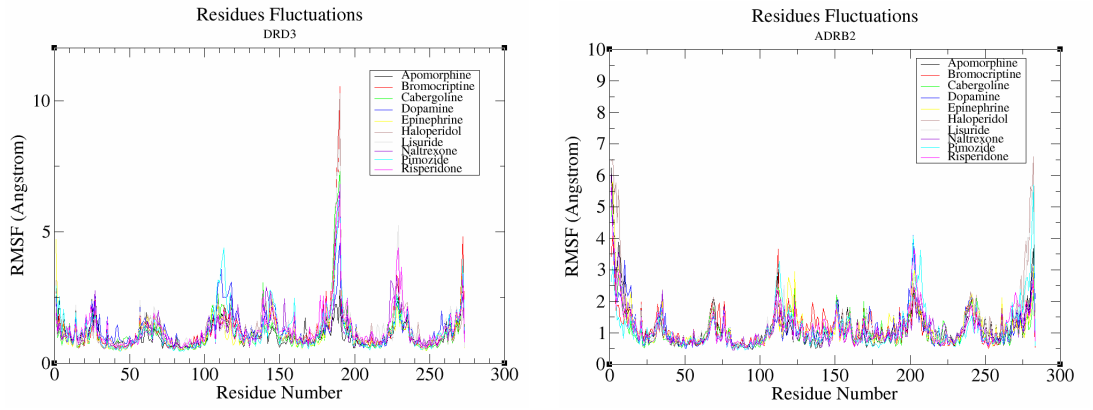


FIGURE 7.10: RMSF fluctuations for all atoms of the receptors showing residue fluctuations during the production MD simulation for the selected drugs complexed with DRD3 and ADRB2 receptors.

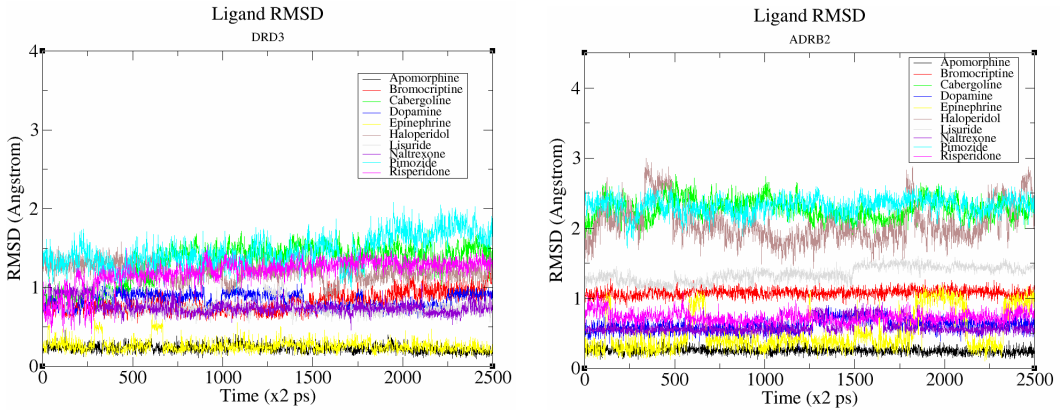
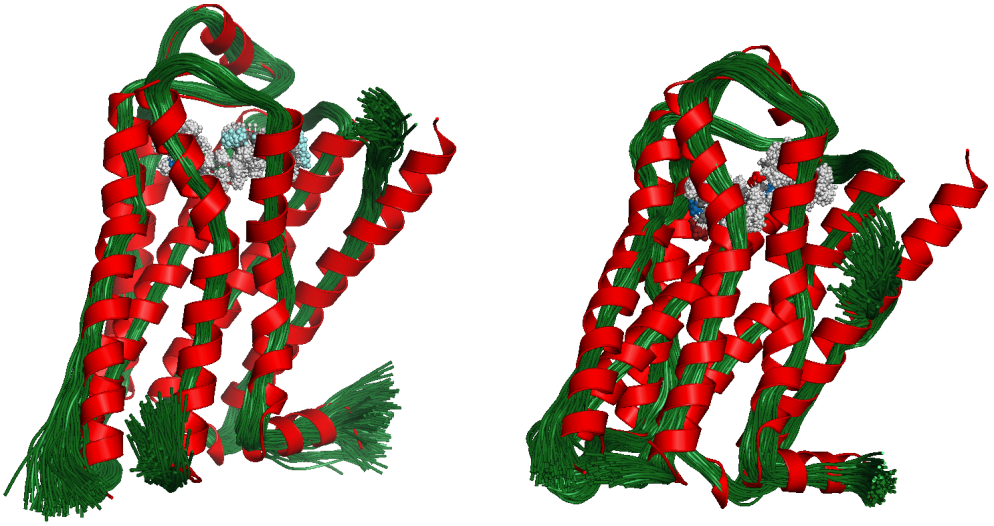


FIGURE 7.11: RMSD plots calculated for the ligand atoms only during the production MD simulation for the selected drugs complexed with DRD3 and ADRB2 receptors.

7.3.3.3 Free Energy of Binding

Binding free energies were estimated for all the selected drug-target complexes. Using MM-PBSA/GBSA based methods to calculate binding free energies, representative snapshots at every 10 ps were extracted from the MD simulation trajectories. The estimated free energy values were compared against experimentally determined activity values for two targets DRD3 and ADRB2 for which the activity data of selected drugs were available. The experimental and calculated binding free energy values (ΔG) are given in Table-7.4.

A positive correlation was observed between the experimental ΔG values and calculated ΔG values for both ADRB2 and DRD3 receptors. The observed correlation is 0.79 and 0.80 for DRD3 determined by both MM-PBSA and MM-GBSA methods respectively, and correlation of 0.54 and 0.50 for ADRB2 using both methods.



(a) Pimozide Complex (b) Bromocriptine Complex

FIGURE 7.12: Comparison of 166 MD snapshots taken at 15ps intervals each from the production MD trajectory, with aligned coordinates of ADRB2 showing fluctuating regions at the extra-cellular end of ADRB2 receptor.

TABLE 7.4: Experimental (exp.) activity values in nM and Δ G values in kcal/-mol for the selected drugs against DRD3 and ADRB2 receptors obtained from KiDB and BindingDB. ΔG (exp.) values are obtained by using Equation 7.1.

	DRD3		**ADRB2**	
Drug Name	**Ki (nM)**	**ΔG (exp.)**	**Ki (nM)**	**ΔG (exp.)**
Apomorphine	36.15	-10.01	10,000	-6.75
Bromocriptine	7.06	-11.19	741	-8.27
Cabergoline	0.79	-12.29	10,000	-6.75
Dopamine	60	-9.75	50,000	-5.77
Epinephrine	-	-	2,065	-7.67
Haloperidol	9.49	-10.89	5,000	-7.16
Lisuride	0.79	-12.29	7.94	-10.93
Pimozide	5.7	-11.13	5,000	-7.16
Risperidone	6.08	-11.90	5,000	-7.16

The reason for low correlation of binding energy with activity values could be the higher affinity values of these drugs for ADRB2 target as compared to DRD3. The correlations of free energy of binding values with experimental ΔG values are plotted in Figure- 7.13. The free energies of binding predicted by MM-PBSA method are quite sensitive to the solute dielectric constant parameter [285]. Using the default dielectric constant (i.e. indi=1.0), the results were not satisfactory in terms of binding energy estimation (Table 7.5). The predicted energies for some drugs were positive and the correlation with experimental ΔG was 0.27 in case of DRD3 and 0.49 for ADRB2. Also there was very low correlation between the energy

Table 7.5: MM-PBSA based estimated ΔG values (kcal/mol) for DRD3 and ADRB2 receptors using default parameter (indi=1.0) for solute dielectric constant.

	DRD3		ADRB2	
Drug Name	**Δ G**	**STD**	**Δ G**	**STD**
Apomorphine	-12.73	5.73	14.16	7.15
Bromocriptine	3.71	5.66	-12.21	5.4
Cabergoline	-8.75	6.07	8.55	6.53
Dopamine	2.17	6.48	9.31	6.12
Haloperidol	-3.7	4.78	-0.79	4.88
Lisuride	-9.77	4.95	-2.96	4.99
Pimozide	-3.75	4.56	-7.87	4.66
Risperidone	-4.7	5.82	11.29	6.08

Table 7.6: MM-PBSA/GBSA based estimated ΔG values (kcal/mol) for DRD3 receptor using solute dielectric constant of 4.0 in PBSA calculations.

	PBSA		GBSA	
Drug Name	**Δ G**	**STD**	**Δ G**	**STD**
Apomorphine	-17.09	2.92	-34.54	2.70
Bromocriptine	-23.11	3.32	-53.02	3.43
Cabergoline	-29.81	3.78	-60.58	4.42
Dopamine	-8.55	2.4	-21.51	2.86
Epinephrine	-13.11	2.9	-31.40	3.13
Haloperidol	-18.94	2.85	-45.54	3.01
Lisuride	-21.00	2.74	-44.13	2.61
Pimozide	-23.02	3.46	-50.93	3.56
Risperidone	-20.08	3.44	-49.05	3.21

Table 7.7: MM-PBSA/GBSA based estimated ΔG values (kcal/mol) for ADRB2 receptor using solute dielectric constant of 4.0 in PBSA calculations.

	PBSA		GBSA	
Drug Name	**Δ G**	**STD**	**Δ G**	**STD**
Apomorphine	-7.35	3.19	-25.62	2.43
Bromocriptine	–36.88	3.93	-74.04	3.48
Cabergoline	-19.61	3.36	-43.80	4.26
Dopamine	-4.59	2.34	-10.25	3.14
Epinephrine	-11.20	2.97	-26.55	4.24
Haloperidol	-7.32	6.32	-15.62	10.44
Lisuride	-18.11	3.76	-37.49	5.26
Pimozide	-22.89	3.06	-55.36	2.99
Risperidone	–21.27	2.89	-44.06	3.12

values determined by both MM-PBSA and MM-GBSA methods. A correlation of 0.12 was observed between PBSA and GBSA for DRD3 and 0.53 for ADRB2.

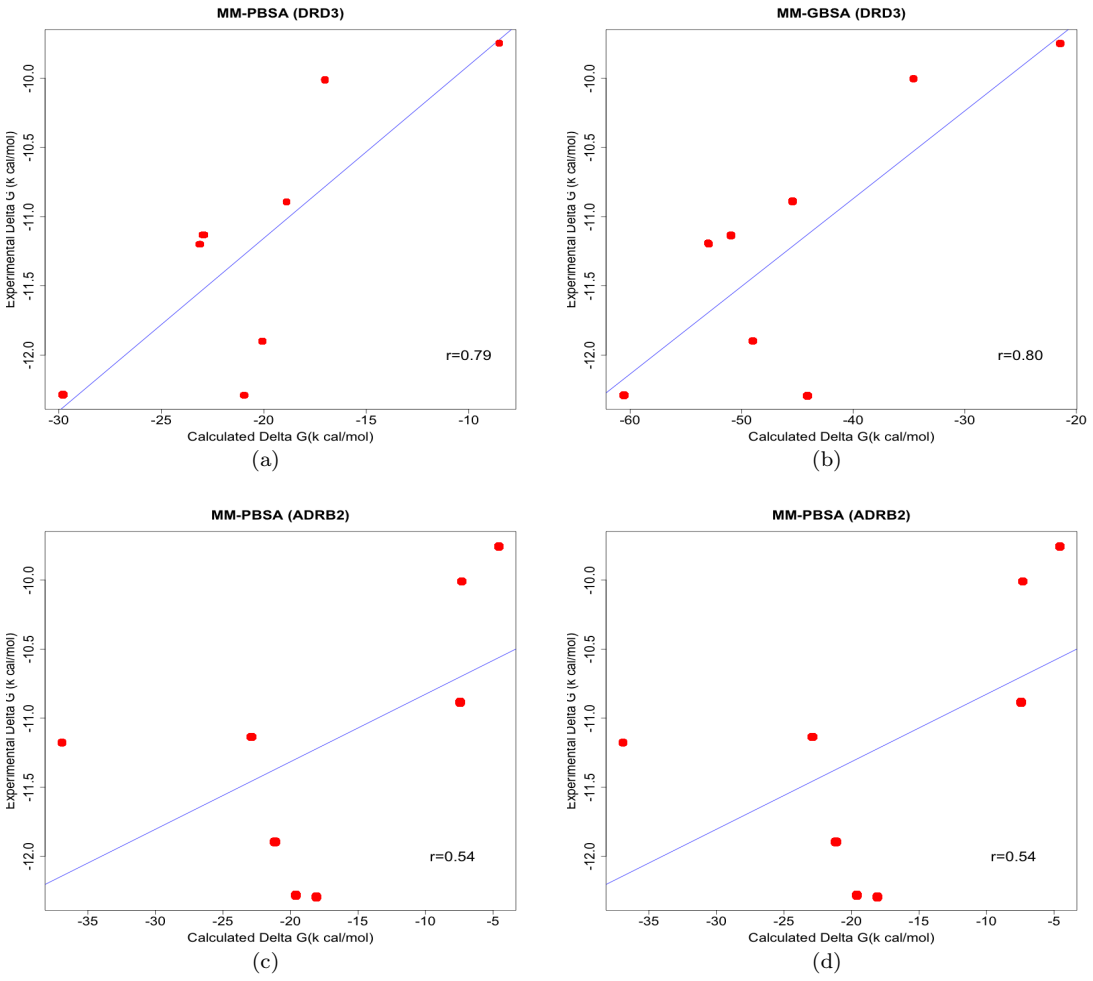


FIGURE 7.13: Correlation plots of experimental Δ G compared to calculated Δ G determined by MM-PBSA and MM-GBSA for selected drugs against DRD3 and ADRB2 receptors.

Therefore, the parameter for solute dielectric constant was increased from 1.0 to 4.0 that improved the results of MM-PBSA for both receptors (Table 7.6 and Table 7.7). Using the higher dielectric constant resulted in agreement for binding free energy estimation between MM-PBSA and MM-GBSA methods for both receptors with a correlation of 0.98 and 0.96 for DRD3 and ADRB2 respectively.

The free energy of binding values were also estimated for the same set of drugs against the remaining GPCR targets: ADORA2A, HRH1 and CXCR4. The drugs were docked and the drug-bound complexes were simulated in the same way as done for the DRD3 and ADRB2 targets. The resulting estimated binding energy values obtained from cross-docking the set of selected drugs against the GPCR targets shows that some drugs bind with high predicted free energy of binding to more than one target as shown in Table-7.8. This indicates that certain drugs have promiscuity and have dual and multiple targets.

Table 7.8: MM-PBSA based calculated ΔG values kcal/mol for 5 GPCR targets.

Drug Name	DRD3	ADRB2	ADORA2A	CXCR4	HRH1
Apomorphine	-17.09	-7.35	-14.99	-6.55	-12.77
Bromocriptine	-23.11	-36.88	**-29.93**	-12.25	ND
Cabergoline	-29.81	-19.61	-13.02	-11.47	-13.48
Dopamine	-8.55	-4.59	-2.44	-6.86	-12.41
Epinephrine	-13.11	-11.20	-5.34	-6.06	-11.51
Haloperidol	-18.94	-7.32	**-20.05**	-10.30	**-24.32**
Lisuride	-21.00	-18.11	-10.43	-14.12	-13.68
Pimozide	-23.02	-22.89	**-22.82**	-11.55	-16.31
Risperidone	-20.08	-21.27	-17.59	-5.86	**-26.94**

ND: No docking predictions.

The results show that drugs such as Bromocriptine, Haloperidol, Pimozide and Risperidone, who are active against DRD3 receptor with affinity less than 10nM (see Table 7.4), can possibly demonstrate activity against a set of secondary targets. These drugs have higher affinities for their primary target DRD3, but additionally, high free energies of binding (highlighted in bold in Table 7.8) were also predicted for the other GPCR targets such as ADORA2A and HRH1. Therefore, these drugs may have polypharmacology and they could exhibit the characteristics of multitarget interacting drugs.

7.3.4 Focused Libraries Enrichement

Top 500 ranked compounds were selected from the compound database on the basis of probability values calculated by the underlying classifier model. Figure-7.14 shows the decision values plot for the database compounds that were distributed around the central decision line at 0. The decision values are calculated by the underlying SVM model that assign class labels to the data points. The compounds falling far from the central decision line (SVM decision boundary) towards the positive scale are more likely to be useful in NDD. Therefore, the top 500 compounds were selected that have a greater probability of being similar to NDD drugs in molecular properties. Similarly, the top 500 compounds were also filtered from the compound database on the basis of similarity scores obtained by applying fingerprint based similarity measure.

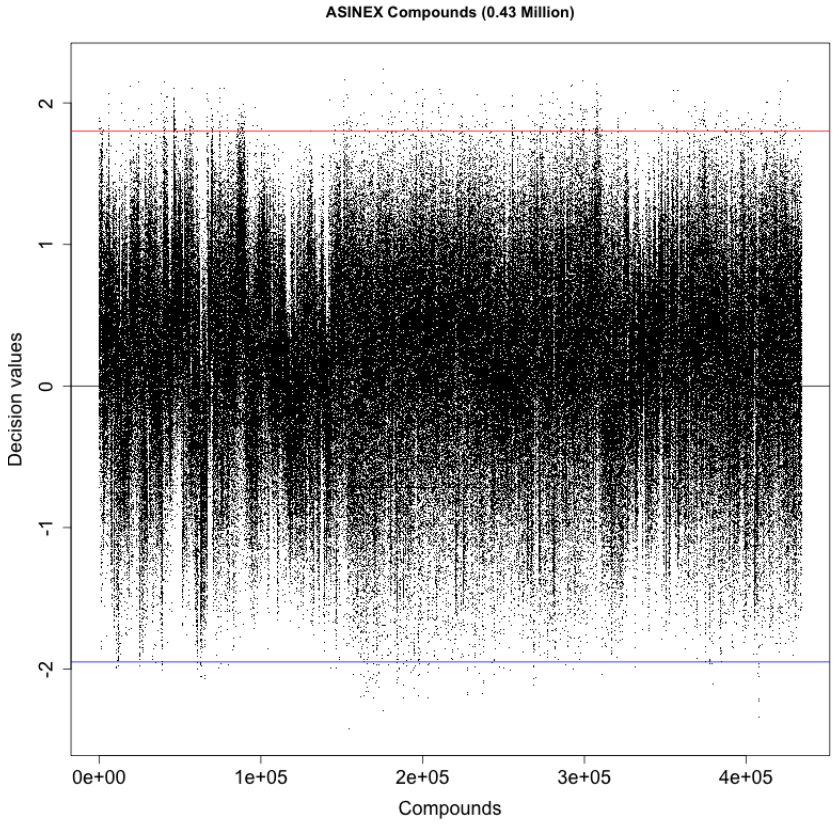


FIGURE 7.14: Ranking and selection of NDD focused library by using machine-learning classification model based on QSAR descriptors. Compounds falling above the line at 0 are predicted as NDD compounds while compounds falling the line below 0 are predicted as non-NDD like compounds. Top 500 ranking compounds are selected that are indicated with red line and the compounds below the blue line are taken as negative samples.

7.3.5 Docking, Pose Ranking and Selection

Docking was used to generate different conformations of a ligand inside the active site pocket of the protein. The focused library was docked against all of the selected targets and for each compound, with the top 10 docked conformations retained for subsequent NNScore pose ranking and filtering. NNScore-based pose ranking and selection was performed because it demonstrated performance in recovering the known active ligands for four of the selected targets that were evaluated for pose ranking. As shown in Figure 7.15, both NNScore and Protein-Ligand Interaction Fingerprint (PLIF) based ZScore performed better in recovering known active ligands as compared to the built-in scoring function of FlexX docking program. NNScore performed better for DRD3 screening and demonstrated early enrichment for ADORA2A screening, while for MAOB and HRH1, the early enrichment achieved with NNSscore was similar to PLIF-based Zscore. Therefore, it is hard

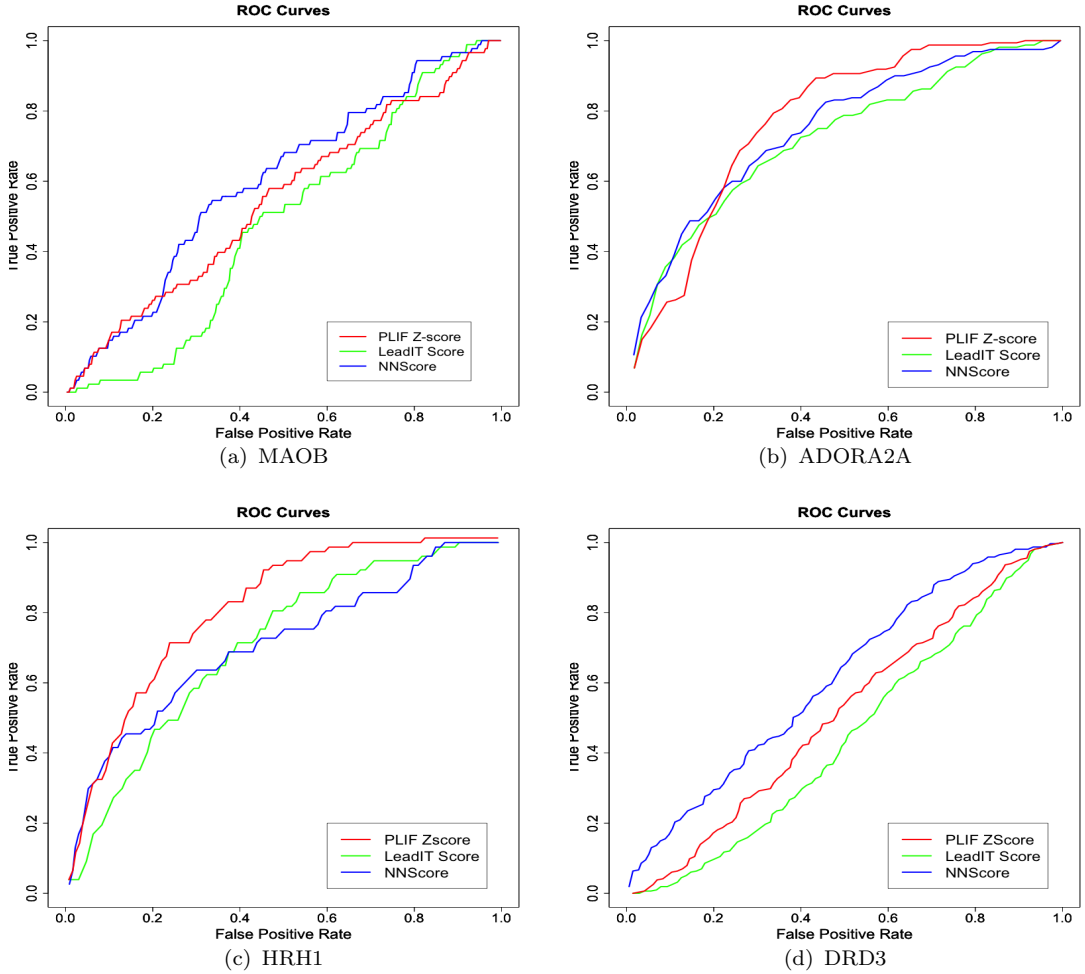
to choose a method on the basis of a screening datasets of a few targets. NNScore was chosen as a method for pose scoring and ranking because its implementation is comparatively simpler than that of PLIF based method. PLIF-based scoring and pose selection method has been described in detail in Chapter 6. However, there are still improvements needed in the scoring functions in general, for pose ranking and selection, because an important point to note is that such performance evaluations should be carried out between experimentally determined true non-binders as decoys and the bio-active molecules.

Ranking among the top 10 conformations and between all compounds of the focused library were obtained for all the targets. Initially, ranks were obtained for all the library compounds against individual targets. Afterwards, sorting was done on the basis of average ranks obtained across all the selected targets. This resulted in a set of 50 compounds that have high average ranks for the targets. The advantage of prioritizing a compound that have high average rank for all targets is to select a potential candidate that may act as a multitarget ligand. A disadvantage in this approach will be down ranking a potential active and target selective compound. Therefore, for the study of targeted polypharmacology, candidates with higher average ranks obtained for all targets will increase the probability of selecting compounds that might exhibit polypharmacology. The selected compounds are later evaluated through more rigorous methods based on MD simulations and subsequent MM-PBSA/GBSA calculations.

7.3.6 MD Simulations of Selected Compounds

Figure-7.16 and Figure 7.17 show 2D structural representations of the compounds that were finally selected for further screening against the selected targets using MD simulation. Postprocessing of the MD trajectories resulted in free energy of binding estimates for all the 50 compounds against the selected targets as shown in Figure-7.18. These results show that there is high positive correlation between the free energy values estimated by MM-PBSA and MM-GBSA methods for all GPCR and enzyme targets. The correlation values between MM-PBSA and MM-GBSA obtained for the targets are given in Table-7.9.

This indicates that for all the targets there is high positive correlation between the binding free energy values determined by both methods. Therefore, both methods can be used in the final ranking and selection of candidate compounds. Studies have shown that MM-PBSA performs better than MM-GBSA in calculating binding free energies[285] and MM-PBSA based binding energy values provide good correlation with experimental activity values [305] when longer MD simulations



(a) MAOB (b) ADORA2A

(c) HRH1 (d) DRD3

FIGURE 7.15: ROC curves showing the performance of different scoring functions in recovering known active ligands from a set of decoy compounds.

TABLE 7.9: Correlation between binding free energy values estimated by both MM-PBSA and MM-GBSA methods for the selected compounds simulated against the selected targets.

Target Protein	**Correlation**
ADRB2	0.88
ADORA2A	0.90
DRD3	0.84
HRH1	0.75
CXCR4	0.90
MAOB	0.81
PDE4A	0.80

are run. However, MM-PBSA calculations are expensive as they use explicit solvent models as compared to the computationally efficient MM-GBSA calculations

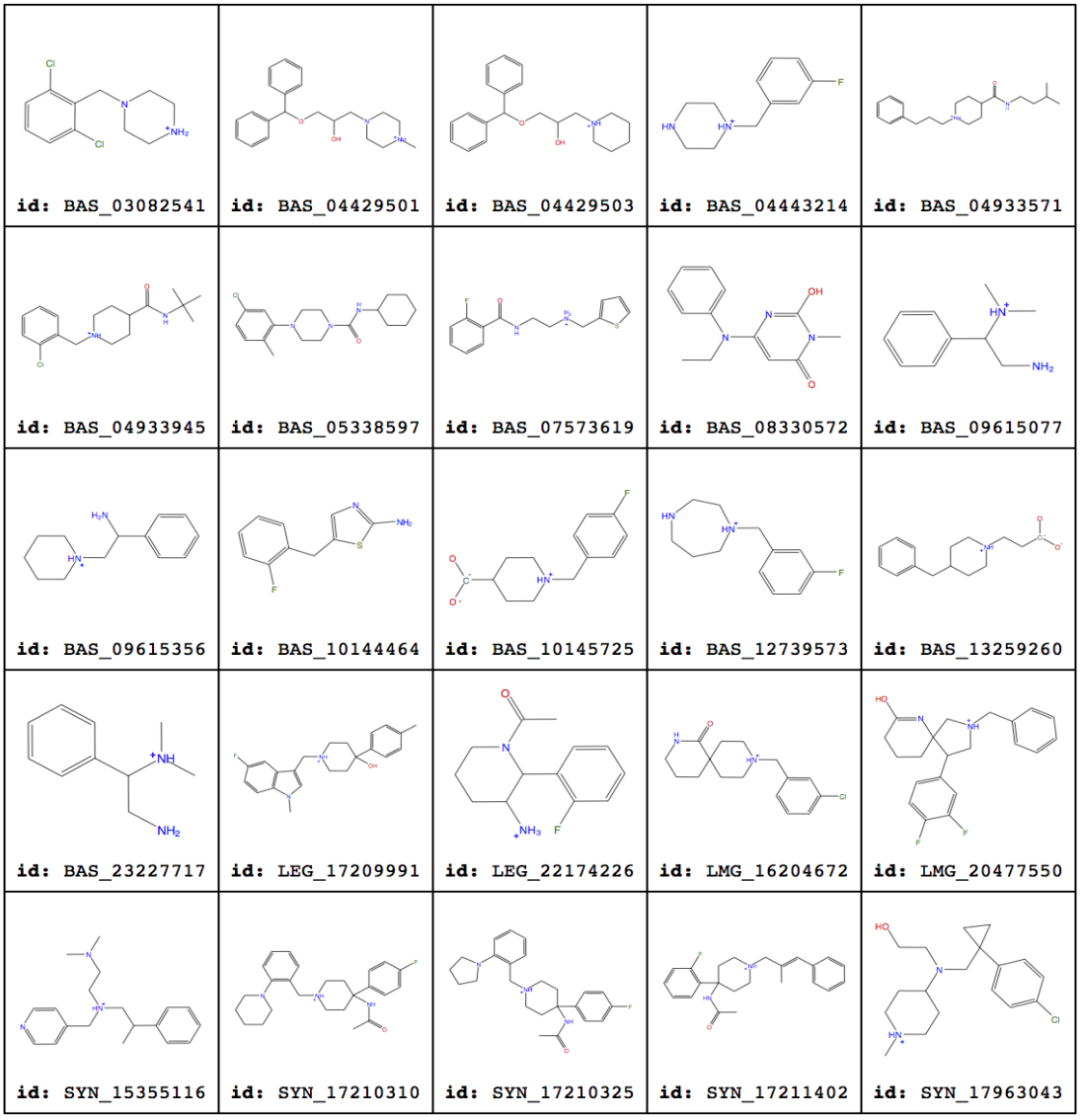


FIGURE 7.16: 2D structural representations of selected compounds with corresponding database IDs.

that use implicit solvation model.

7.3.7 Polypharmacology

The estimated binding energy values obtained for the selected compounds against multiple target proteins showed that some of the compounds are predicted to be more promiscuous than others. Therefore, it is assumed that compounds with greater promiscuity will tend to interact with a similar strength of binding energy with multiple target proteins that have binding site similarities. Likewise, the compounds that do not exhibit possible polypharmacological interaction behavior will not bind favorably to multiple targets. In other words, the same ligand should

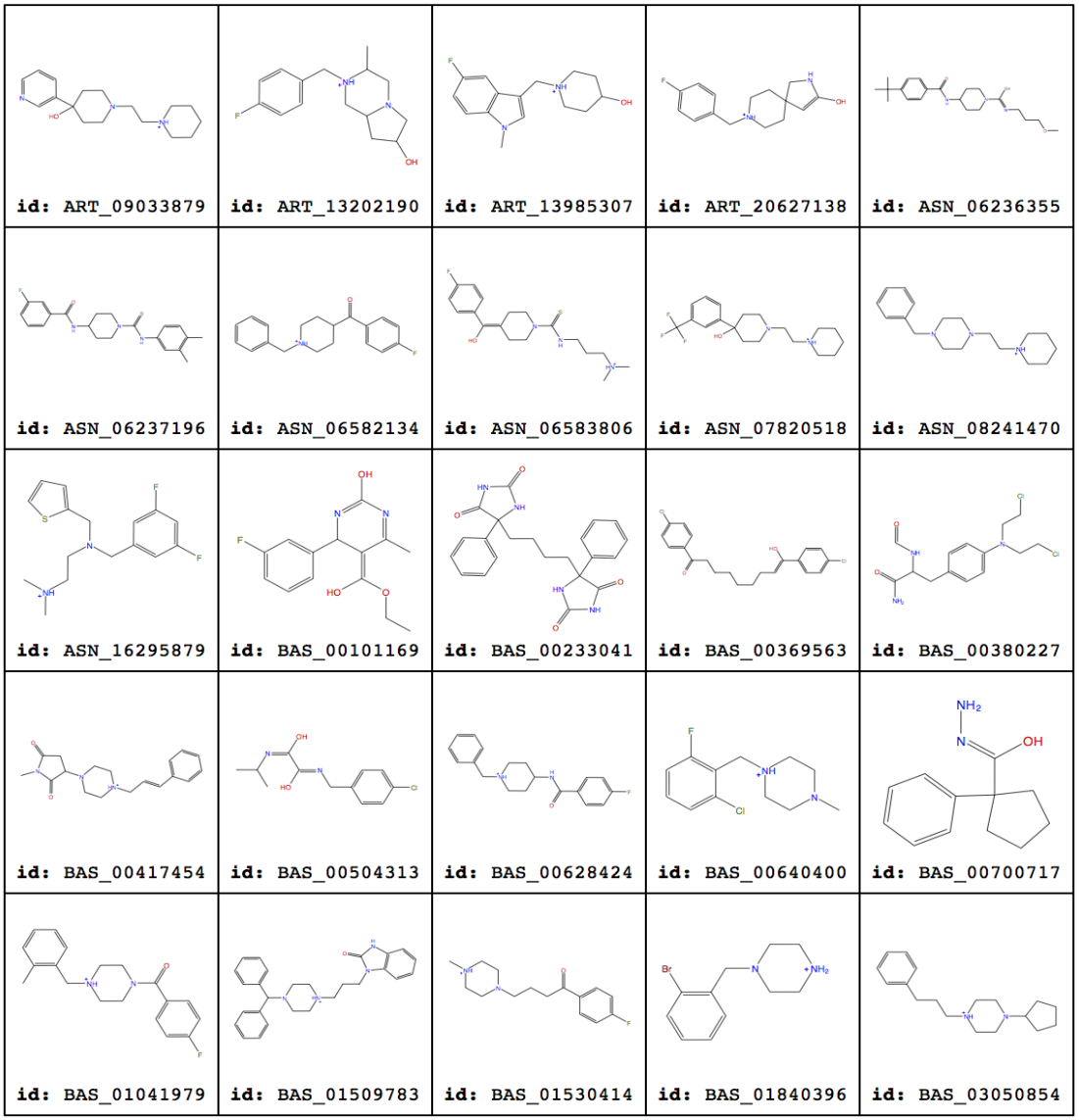


FIGURE 7.17: 2D structural representations of selected compounds with corresponding database IDs.

bind with similar binding energy to multiple pockets that have high similar pockets and the binding energy should vary in a dissimilar pocket. To test the hypothesis that the compounds tend to interact with similar binding energies with multiple targets having binding pocket similarities, a comparison is made between the correlation of binding energy values of the same compounds with different targets, and the pocket similarity values between different targets. A correlation matrix is generated from the correlation values of the same compounds with multiple targets that is compared with the pocket similarity matrix of the selected targets.

Figure-7.19 (A and B) shows the pocket similarity matrix visualized as heatmap that indicates the pocket similarity range for 5 GPCRs and 2 enzyme targets.

From this figure, it is clear that the GPCR targets have higher binding pocket similarities with each other as compared to the two enzyme targets. Such similarities are expected because the overall fold of these receptors is conserved [306] and sharing similar ligands my reflect the conservation of identical ligands between unrelated receptors [307]. Figure (C and D) show the heatmaps generated from the correlation of binding energy values computed for the selected compounds against different targets through MM-PBSA and MM-GBSA methods respectively. It can be observed that the correlation is slightly higher for some GPCR targets as compared to the enzyme targets. Therefore, it can be assumed that the selected compounds may exhibit polypharmacological behavior towards the GPCR targets more than the enzymes, behavior that can be confirmed by performing experimental validations.

7.3.8 Comparative QSAR Modeling of Novel Candidate Compounds

The selected compounds that were predicted as potential candidates for use in NDD disorders were compared with a set of FDA approved drugs for various NDD disorders. The objective of this comparison was to find common features between the newly predicted candidates and established NDD drugs. To test the functionality of the QSRA algorithms, a set of compounds which were predicted as non-NDD like compounds by the NDD classifier model, was used as a negative control. QSAR-based predictions for interacting targets, pathways and involvement in disease indications were made for the three sets of compounds/drugs using QSAR-based algorithms available in MetaDrug [175]. The web interface of MetaDrug was used to upload the three data sets. Tanimoto similarity cut-off of 0.7 was used. For each data set, comprehensive reports were generated, which were parsed to extract the following three types of predictions:

a Target predictions

b Pathways predictions based on common targets

c Disease predictions

Table 7.10 lists a set of known FDA approved drugs used for the treatment of four NDD disorders (Alzheimer's, Parkinson's diseases, Epilepsy and Multiple Sclerosis) that are used as a reference set for the comparisons of predicted NDD+ and NDD- compounds. Figure 7.20 shows the results of predicted targets for known drugs, predicted NDD+ and predicted NDD- drugs. It was observed that similar

targets were predicted for the NDD+ compounds that are interacting with known drugs. Moreover, there were almost no targets predicted for the NDD- compounds. Similarly, there were also similarities at pathways level and disease indications predicted based on similar interacting targets for the known drugs as well as for the predicted NDD compounds. The predicted list of targets includes also those GPCR targets against which the current virtual screening was performed, the predicted pathways includes the important neurophysiological processes and the predicted disease indications include Alzheimer's and Parkinson's diseases. These analyses provide another basis for the relevance and potential of the selected compounds for use in NDD indications.

7.3.9 Candidates Selection and Identificaiton of Multitarget Drug

Finally, a set of 14 candidate hits are selected for further experimental validation. The candidate compounds are selected on the basis of binding energy values (i.e. calculated with both MM-PBSA and MM-GBSA methods) that are average ranked for the selected targets. As shown in Figure 7.18, the best candidate compounds are selected that have higher binding energy values for multiple targets. The candidate compounds are currently evaluated experimentally for activities against the selected targets.

The first hit in the selected candidates list is identified to be a known antiasthmatic and antiallergic drug, Oxatomide. Oxatomide is orally active histamine H1 antagonist, chemically related to cinnarizine class [308], which is used for the treatment of allergy and other related indication areas. Our results have shown higher predicted binding energy for Oxatomide (BAS-01509783 in Figure 7.17 and Figure 7.18) against 4 out of 7 targets. Our targeted polypharmacology approach has predicted this drug as a multitarget drug, because experimental activities for this drug against multiple targets have been reported. For its primary target, HRH1 receptor, IC50 of 9 nM [309], for DRD3 receptor, Ki of 62 nM [310] and for CACNA1H (Voltage-dependent T-type calcium channel subunit alpha-1H), an EC50 of 5150 nM [311] have been reported. This shows that for two targets, the predictions for this compound turned out to be correct and for the rest of the targets it remains to be validated experimentally. Therefore, it is more likely that some of the candidate compounds might also exhibit multitarget activities that will demonstrate polypharmacology, after the activities are reproduced in *in-vitro* experiments.

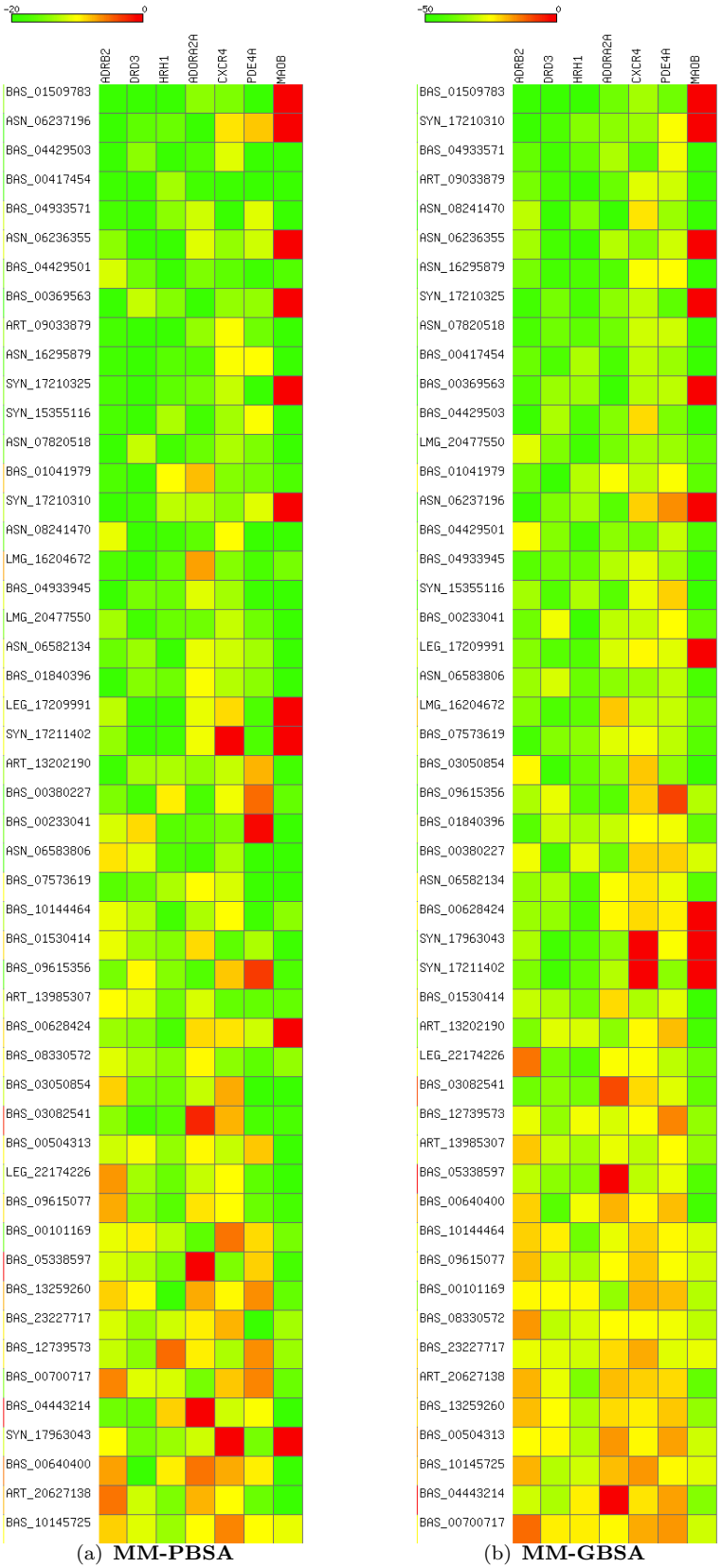
TABLE 7.10: List of FDA approved drugs and their use in various NDD indications. Drug ids are pointing to the DrugBank, the TTD or STITCH databases.

Drug ID	**Name**	**Indication area**
DB00674	Galantamine	Alzheimer's Disease
DB00843	Donepezil	Alzheimer's Disease
DB01043	Memantine	Alzheimer's Disease
DB00989	Rivastigmine	Alzheimer's Disease
DB00382	Tacrine	Alzheimer's Disease
DB06262	Droxidopa	Parkinson's Disease
DB00190	Carbidopa	Parkinson's Disease
DB00494	Entacapone	Parkinson's Disease
DB01367	Rasagiline	Parkinson's Disease
DB00714	Apomorphine	Parkinson's Disease
DB00915	Amantadine	Parkinson's Disease
DB00387	Procyclidine	Parkinson's Disease
DB01200	Bromocriptine	Parkinson's Disease
DB01235	Levodopa	Parkinson's Disease
DB00413	Pramipexole	Parkinson's Disease
DB05271	Rotigotine	Parkinson's Disease
DB01037	Selegiline	Parkinson's Disease
DCL000972	Safinamide	Parkinson's Disease
DB00268	Ropinirole	Epilepsy
CID000000444	Bupropion	Epilepsy
CID000002771	Citalopram	Epilepsy
CID000004763	Phenobarbital	Epilepsy
DCL000791	Eslicarbazepine acetate	Epilepsy
DCL000982	Seletracetam	Epilepsy
DCL000730	Brivaracetam	Epilepsy
CID000004192	Midazolam	Epilepsy
CID000001986	Acetazolamide	Epilepsy
DB00564	Carbamazepine	Epilepsy
CID000003016	Diazepam	Epilepsy
DB00593	Ethosuximide	Epilepsy
DB00949	Felbamate	Epilepsy
DB01320	Fosphenytoin	Epilepsy
DB00996	Gabapentin	Epilepsy
DCL000858	Lacosamide	Epilepsy
DB01202	Levetiracetam	Epilepsy
DB00776	Oxcarbazepine	Epilepsy
DB00252	Phenytoin	Epilepsy
DB00794	Primidone	Epilepsy
DAP001516	Rufinamide	Epilepsy
DB00906	Tiagabine	Epilepsy
DB01080	Vigabatrin	Epilepsy
DB00909	Zonisamide	Epilepsy
CID000002265	Azathioprine	Epilepsy
DB00181	Baclofen	Multiple Sclerosis
DB00108	Natalizumab	Multiple Sclerosis
CID005360515	Naltrexone	Seizure

7.4 Conclusions

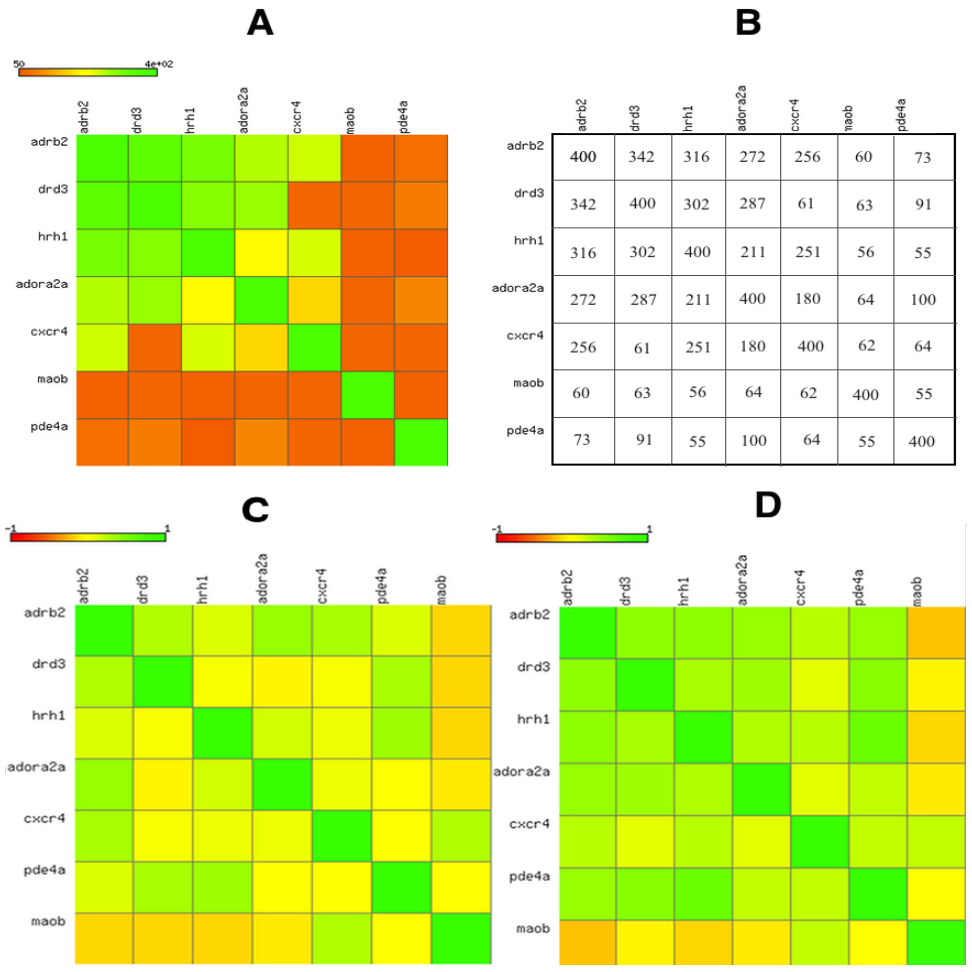
When dealing with complex diseases such as neurodegenerative diseases, knowledge and data from different sources should be obtained because an individual data

source is not always complete. This includes data for both drugs and their targets. However, the key challenges arise from the lack of sufficient data at the desired resolution. In this chapter, we have explored polypharmacology through various aspects by combining network pharmacology with ligand-based and structure-based design methods. An NDD focused ligand library was designed by using our NDD classifier model and similarity searching. Protein targets for the study of targeted polypharmacology were selected by using network pharmacology approach. Using classifier model and similarity search method for our ligand-based approach provided highly enriched and focused library of compounds. This focused library can increase the chances of hit finding as they have similar physiochemical properties to known NDD drugs. Another important aspect is to limit the number of database compounds that are screened against multiple targets for the study of polypharmacology because of the heavy demand for computational resources. For our structure-based design approach, highly prioritized set of protein targets that have clear therapeutic roles in the disease mechanism were selected. Such highly focused set of important therapeutic targets can be thoroughly studied for the binding interactions at both *in-silico* and *in-vitro* environments. If resources are available, polypharmacology by high throughput MD simulations should be performed on as many drug targets as possible that has the potential to reveal interesting predictions of activity of new candidate compounds. This would ultimately facilitate the selection of potential multi-target inhibitors that can be further investigated for the desired polypharmacological profiles and efficacies.



(a) **MM-PBSA** (b) **MM-GBSA**

FIGURE 7.18: Heatmap of free energies of binding values calculated for selected 50 compounds simulated against 7 targets using MM-PBSA and MM-GBSA based methods.



	adrb2	drd3	hrh1	adora2a	cxcr4	maob	pde4a
adrb2	400	342	316	272	256	60	73
drd3	342	400	302	287	61	63	91
hrh1	316	302	400	211	251	56	55
adora2a	272	287	211	400	180	64	100
cxcr4	256	61	251	180	400	62	64
maob	60	63	56	64	62	400	55
pde4a	73	91	55	100	64	55	400

FIGURE 7.19: Speculating polypharmacology by pocket similarity. A) Heatmap of pocket similarity matrix. B) Pocket similarity scores. C) Correlation matrix of MM-PBSA-based derived free energy values of selected 50 compounds between multiple targets. D) Correlation matrix of MM-GBSA-based derived free energy values of selected 50 compounds between multiple targets.

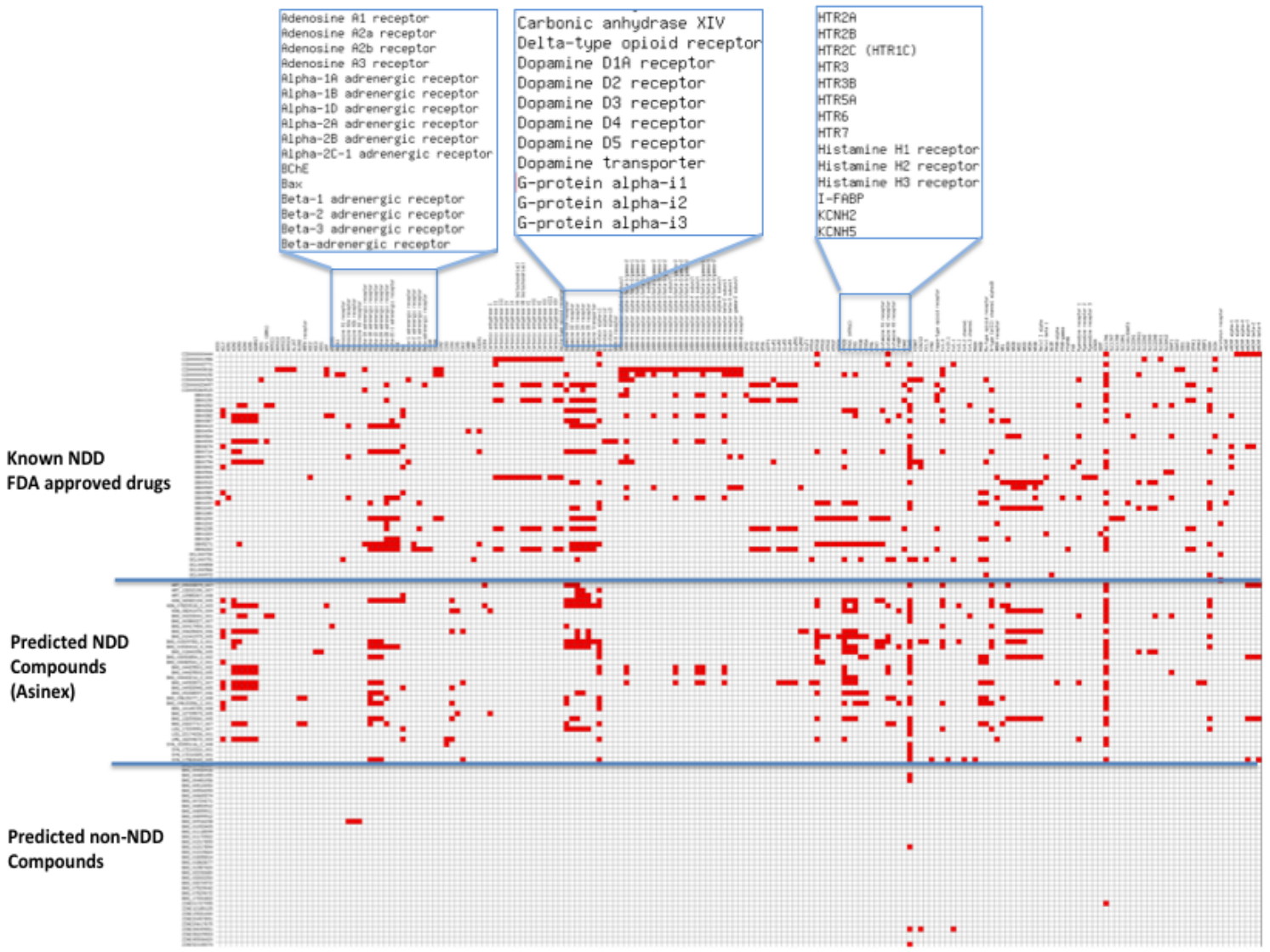


FIGURE 7.20: Heatmap of drug-target interactions predicted by MetaDrug algorithms for three groups of compounds. Compounds are given in rows and targets are given in columns. Black lines divide the compounds groups into FDA approved drugs, predicted NDD compounds and predicted non-NDD compounds. Important target interactions are zoomed into the boxes above the columns.

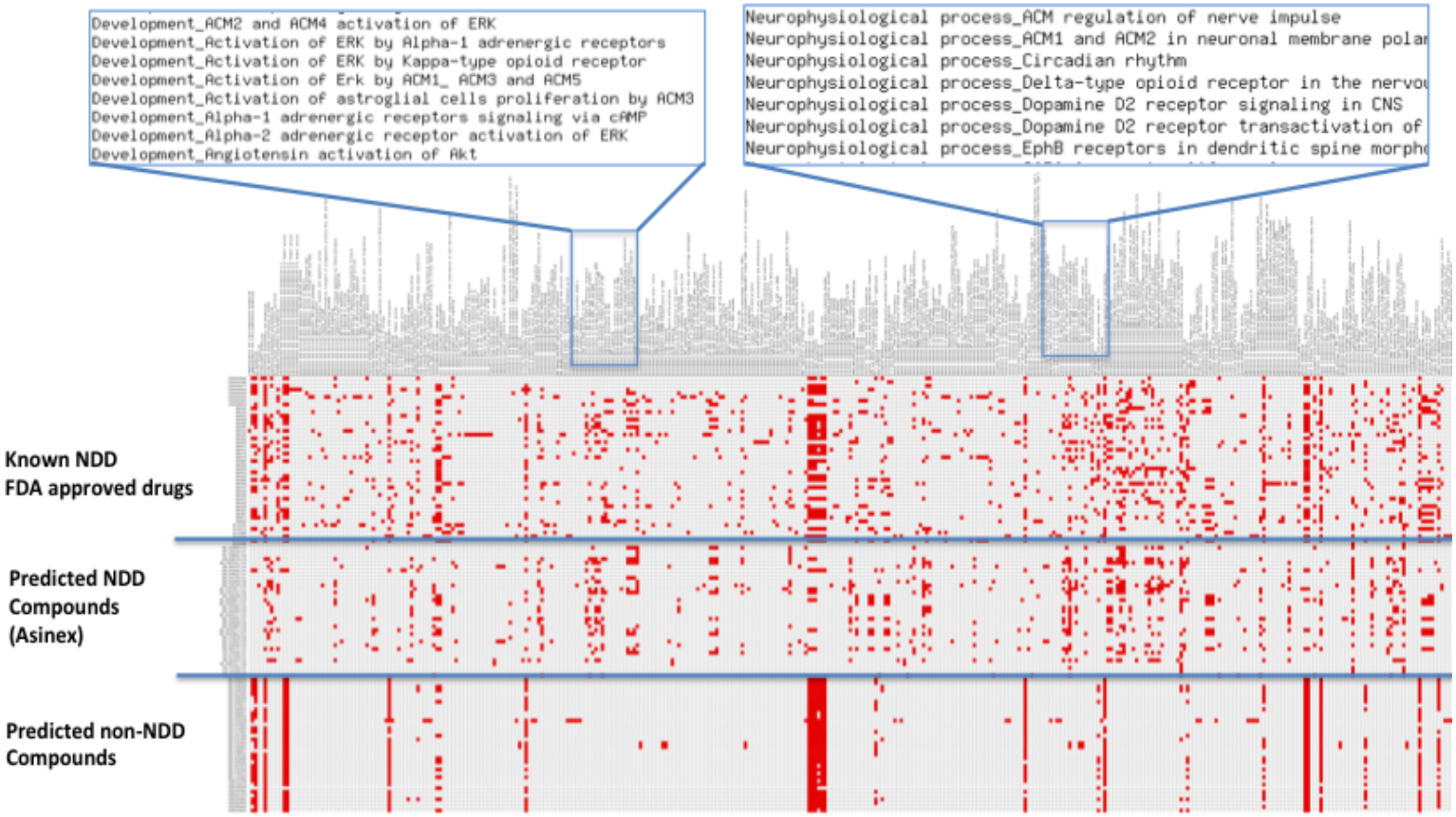


FIGURE 7.21: Heatmap of drug-pathways involvement predicted by MetaDrug algorithms for three groups of compounds. Compounds are given in rows and pathways are given in columns. Black lines divide the compounds groups into FDA approved drugs, predicted NDD compounds and predicted non-NDD compounds. Some pathways are highlighted into the boxes above the columns.

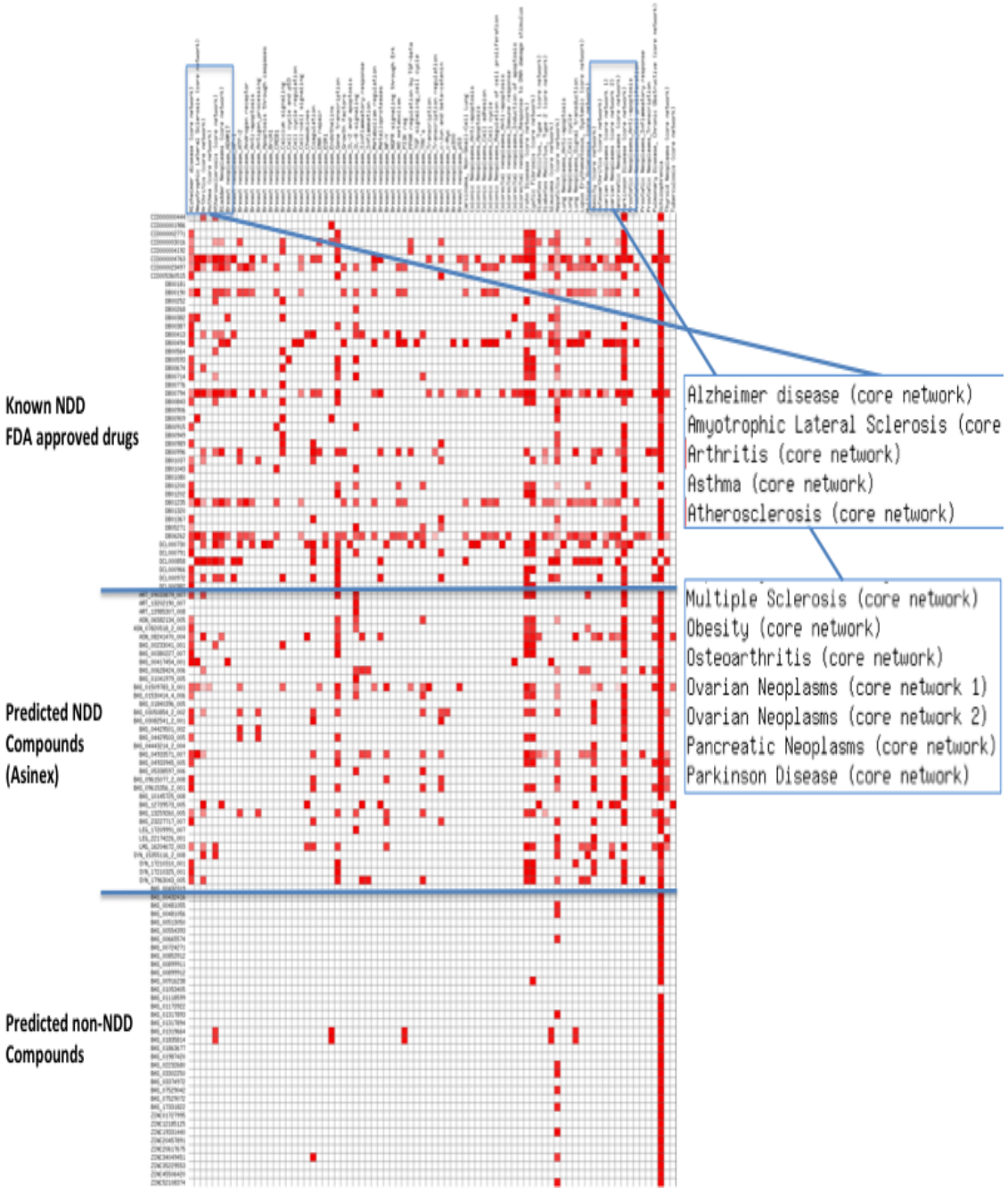


FIGURE 7.22: Heatmap of drug-disease associations predicted by MetaDrug algorithms for three groups of compounds. Compounds are given in rows and diseases are given in columns. Black lines divide the compounds groups into FDA approved drugs, predicted NDD compounds and predicted non-NDD compounds. Some diseases are highlighted into the boxes above the columns.

Chapter 8

Conclusions and Future Perspectives

8.1 Conclusions

Studying integrative systems approaches towards brain pharmacology and polypharmacology starts with the collecting and interlinking of exiting data from available resources that are disintegrated and heterogeneous. This collection of dispersed pieces of data and interlinking them requires the design of a knowledge base system that enables the proper organization of the collected data. The design of the Human Brain Pharmacome is a first effort in this direction to build a comprehensive knowledge base framework for the brain pharmacology that represents and integrates therapeutic data around human brain disorders. Major components of this knowledge base system are the known drugs used in the treatment of major brain disorders and their interacting protein targets. For all of the drugs and their protein targets, the biological and chemical properties and other relevant features are also collected. This knowledge base system should be as complete and comprehensive as possible. The HBP design and foundation is developed during the research work in this thesis, which is then used to study targeted polypharmacology. There is no stand alone knowledge base publicly available that contains all the required data regarding our target area in a single place. To build a complete knowledge base system, it becomes necessary to combine multiple resources, capture the required data, organize and interlink the data and efficiently store them at one place in a well structured format. The HBP pharmacome in this research work, has been used as a central repository that enables the exploration of integrative systems approaches towards targeted polypharmacology. Using network pharmacology approaches, the HBP is analyzed in such a way that enables us to

generate a system level view of a specific knowledge that exists in this knowledge base. The underlying drug-target network in the HBP pharmacome is very large and complex. A Cytoscape plugin, referred to as "Pharmacome Tools", has been developed to filter the large networks generated from the HBP data and to reduce their complexity. The plugin has useful features to predict new drug-target associations that could be investigated for the prediction of off-target effects and drug repurposing. Another useful function of this plugin is to select a set of protein targets by prioritizing them for the study of targeted polypharmacology. Similarly, drugs used for the treatment of neurodegenerative disorders (NDD) are investigated collectively to identify the unique properties of these drugs, build machine learning models over them and then apply the model to characterize new potential compounds for use in NDD. The classifier models also demonstrate the advantage of these methods over the conventional similarity based methods. We have effectively used the classifier model to design a highly enriched and NDD-focused library of compounds from a large compound database. This approach of characterizing novel candidate compounds also has the ability to repurpose existing drugs. Using this approach, some drugs that are used in non-NDD indications were predicted to be potential candidates for testing in various CNS disorders. The same approach was also applied and evaluated on GPCR class of protein targets to identify GPCR antagonists from non-GPCR antagonists and it also revealed a clear potential of drug repositioning.

Polypharmacology has emerged as a new paradigm of drug discovery in the recent years, and has led to a paradigm shift from the classical drug discovery approaches. In polypharmacology, a single drug binds to multiple therapeutic targets that contribute to overall effectiveness of a treatment. Various approaches were studied during this thesis to explore polypharmacology in the existing drugs that are used for major brain disorders, particularly in the indication areas of NDD. With the goal of discovering new candidate drugs through the use of existing knowledge concerning known drugs, new potential candidates with polypharmacological characteristics were discovered and proposed for further investigation. Some new indication areas different than the primary indication area were also proposed for existing drugs.

Studying polypharmacology demands for a great amount of resources. *In-silico* pharmacology approaches that combine ligand-based and structure-based tools on an array of known drugs and protein targets need substantial time and resources. Computational approaches that are applied in this study of polypharmacology

include network pharmacology based multiple target selection, machine learning-based compounds selection, molecular docking followed by automated pose selection and, afterwards, molecular dynamics simulations to evaluate protein-ligand interactions. Selection of favorable binding modes for a compound inside multiple protein targets is facilitated by automated pose ranking and selection by using external scoring function. A set of 50 compounds was filtered for further study that had favorable binding interactions with 7 different targets including 5 GPCRs and 2 enzymes. Binding energy estimates for the compounds against all targets enabled the ranking of each compound against each target. Finally, a small set of highly promising candidate hits were identified that have the possibility of being multi-target inhibitors. These candidate hits are taken forward for experimental validation.

The first hit in the selected candidates list, Oxatomide, is identified as a known antiasthmatic and antiallergic drug, which is an orally active histamine H1 antagonist and is used for the treatment of allergy and other related indication areas. Our targeted polypharmacology approach has identified this drug as a potential multi-target drug. Experimental activities for this drug against other targets have also been reported. This indicates that the predictions for this compound to be active against multiple targets are supported by independent evidences. For other targets in our set of selected targets, the polypharmacology potential remains to be validated experimentally. Therefore, it is more likely that some of the candidate compounds might also exhibit dual and multi-target activities that will demonstrate polypharmacology, after the activities are reproduced in *in-vitro* experiments. Similarly, some known antibiotic drugs were also predicted for their potential use in neurodegenerative and other CNS related disorders.

8.2 Future Perspectives

The integrative systems approaches towards targeted polypharmacology in the human brain, studied in this research work, have demonstrated their usefulness. The successful identification of a known drug as a multi-target agent indicates that these approaches we have undertaken can deliver valid outcome. If more of the predicted multi-target compounds are validated in the experimental screening, this will strengthen our hypothesis of polypharmacology. We believe that there remain challenges and room for improvement at each step of this polypharmacology workflow. An up-to-date HBP pharmacome has the potential to play a significant role in providing the foundation for a wide spectrum of approaches in computational brain pharmacology. Therefore, an automated update mechanism

is very crucial for frequently updating the HBP with new knowledge that arrives in the subject domain. The HBP design allows for extension of the existing data and can accommodate multidimensional chemical and biological data that can be very useful for many purposes relevant to the pharmaceutical industry.
The HBP will continue to extend by adding and integrating more knowledge from additional sources. The HBP design allows for the application of mining strategies that will help grow the pharmacome by retrieving the hidden knowledge from biomedical literature, patent documents and patients health records. The HBP boundaries can be crossed to collect additional data for all FDA approved drugs for research on drug repositioning. Mining strategies to collect data from literature about drugs intervention at pathways level and integrating gene expression profiles of the interacting protein targets can be used for the identification of true mode of action of drugs. Addition of tissue specific protein-protein interaction data for the disease and normal states and linking them with phenotypic associations at the organ and organism levels will enable the pharmacome to be used for *in-silico* target identification strategies. With the availability of growing compute power and resources, HBP can support large scale virtual screening platforms by providing ample data for pharmacology applications to reveal new and interesting findings. The HBP, as a unique pharmacology knowledge base, continues to develop along the future route for understanding complex biomolecular interactions occurring in the brain, thus providing opportunities for developing effective treatment strategies for complex brain diseases.

References

[1] Donald Silberberg. The high impact of neurologic disorders in developing countries: The struggle for global recognition. *Neurology*, 77(3):307–308, 2011. doi: 10.1212/WNL.0b013e3182285da9. URL http://www.neurology.org/content/77/3/307.short.

[2] A. Valdes-Sosa Pedro. Coping with brain disorders using neurotechnology. *The Malaysian Journal of Medical Sciences*, 9:1–3, 2012.

[3] Farrah J. Mateen. Neurocritical care in developing countries. *Neurocritical Care*, 15(3):593–598, 2011. ISSN 1541-6933. doi: 10.1007/s12028-011-9623-7. URL http://dx.doi.org/10.1007/s12028-011-9623-7.

[4] Donald Silberberg. Brain disorders where resources are scarce: The unfinished agenda. *Neurology*, 79(12):1285–1287, 2012. doi: 10.1212/WNL.0b013e31826aad6d. URL http://www.neurology.org/content/79/12/1285.short.

[5] Alice B. Smith, James G. Smirniotopoulos, and Elisabeth J. Rushing. Central nervous system infections associated with human immunodeficiency virus infection: Radiologic-pathologic correlation. *Radiographics*, 28(7):2033–2058, 2008. doi: 10.1148/rg.287085135. URL http://radiographics.rsna.org/content/28/7/2033.abstract.

[6] Esther T. Stoeckli. What does the developing brain tell us about neural diseases? *European Journal of Neuroscience*, 35(12):1811–1817, 2012. ISSN 1460-9568. doi: 10.1111/j.1460-9568.2012.08171.x. URL http://dx.doi.org/10.1111/j.1460-9568.2012.08171.x.

[7] D.F. Boland and M. Stacy. The economic and quality of life burden associated with parkinson's disease: a focus on symptoms. *The American Journal of Managed Care*, 18:168–175, 2012.

[8] J. Olesen, A. Gustavsson, M. Svensson, H.U. Wittchen, B. Jönsson, on behalf of the CDBE2010 study group, and the European Brain Council. The economic cost of brain disorders in europe. *European Journal of Neurology*, 19(1):155–162, 2012. ISSN 1468-1331. doi: 10.1111/j.1468-1331.2011.03590.x. URL http://dx.doi.org/10.1111/j.1468-1331.2011.03590.x.

[9] M. Miret, J. L. Ayuso-Mateos, J. Sanchez-Moreno, and E. Vieta. Depressive disorders and suicide: Epidemiology, risk factors, and burden. *Neuroscience & Biobehavioral Reviews*, 2013. ISSN 0149-7634. doi: 10.1016/j.neubiorev.2013.01.008. URL http://www.sciencedirect.com/science/article/pii/S0149763413000092.

[10] Daniel P. Kennedy and Ralph Adolphs. The social brain in psychiatric and neurological disorders. *Trends in Cognitive Sciences*, 16(11):559–572, 2012. ISSN 1364-6613. doi: 10.1016/j.tics.2012.09.006. URL http://www.sciencedirect.com/science/article/pii/S1364661312002227.

[11] David B. Teplow. Preface. In *Molecular Biology of Neurodegenerative Diseases*, volume 107 of *Progress in Molecular Biology and Translational Science*, pages xiii – xiv. Academic Press, 2012. doi: 10.1016/B978-0-12-385883-2.00014-X. URL http://www.sciencedirect.com/science/article/pii/B978012385883200014X.

[12] Hui-Ming Gao and Jau-Shyong Hong. Why neurodegenerative diseases are progressive: uncontrolled inflammation drives disease progression. *Trends in Immunology*, 29(8):357–365, 2008. ISSN 1471-4906. doi: 10.1016/j.it.2008.05.002. URL http://www.sciencedirect.com/science/article/pii/S1471490608001567.

[13] Carlo Bertoni-Freddari, Patrizia Fattoretti, Tiziana Casoli, Giuseppina Di Stefano, Belinda Giorgetti, and Marta Balietti. Brain aging: The zinc connection. *Experimental Gerontology*, 43(5):389 – 393, 2008. ISSN 0531-5565. doi: 10.1016/j.exger.2007.11.001. URL http://www.sciencedirect.com/science/article/pii/S0531556507002690.

[14] R A Knight and A Verkhratsky. Neurodegenerative diseases: failures in brain connectivity[quest]. *Cell Death & Differentiation*, 17(7):1069–1070, July 2010. ISSN 1350-9047. URL http://dx.doi.org/10.1038/cdd.2010.23.

[15] Alzheimer's Association. Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 9(2):208–245, 2013. ISSN 1552-5260. doi: 10.1016/j.jalz.2012.02.001. URL http://www.sciencedirect.com/science/article/pii/S1552526012000325.

[16] Stephen Todd, Stephen Barr, Mark Roberts, and A Peter Passmore. Survival in dementia and predictors of mortality: a review. *International Journal of Geriatric Psychiatry*, 2013. ISSN 1099-1166. doi: 10.1002/gps.3946. URL http://dx.doi.org/10.1002/gps.3946.

[17] Daniel Weintraub, Cynthia L. Comella, and Stacy Horn. Parkinson's disease–part 1: Pathophysiology, symptoms, burden, diagnosis, and assessment. *American Journal of Managed Care*, 14:40–48, 2008.

[18] Patrick Hickey and Mark Stacy. Adenosine A2A antagonists in Parkinson's Disease: What's next? *Current Neurology and Neuroscience Reports*, 12

(4):376–385, 2012. ISSN 1528-4042. doi: 10.1007/s11910-012-0279-2. URL http://dx.doi.org/10.1007/s11910-012-0279-2.

[19] Claire Henchcliffe and W. Lawrence Severt. Disease modification in Parkinson's disease. *Drugs & Aging*, 28(8):605–615, 2011. ISSN 1170-229X. doi: 10.2165/11591320-000000000-00000. URL http://dx.doi.org/10.2165/11591320-000000000-00000.

[20] David J. Pedrosa and Lars Timmermann. Review: management of Parkinson's disease. *Neuropsychiatric Disease and Treatment*, 9:321–340, 2013.

[21] Masoud Etemadifar, Seyed-Hossein Abtahi, Mojtaba Akbari, and Amir-Hadi Maghzi. Multiple Sclerosis and Amyotrophic Lateral Sclerosis: is there a link? *Multiple Sclerosis Journal*, 18(6):902–904, 2012. doi: 10.1177/1352458511427719. URL http://msj.sagepub.com/content/18/6/902.abstract.

[22] Ali Aamer Habib and Hiroshi Mitsumoto. Emerging drugs for amyotrophic lateral sclerosis. *Expert Opinion on Emerging Drugs*, 16(3):537–558, 2011. doi: 10.1517/14728214.2011.604312. URL http://informahealthcare.com/doi/abs/10.1517/14728214.2011.604312. PMID: 21806316.

[23] Vijayshree Yadav and Dennis Bourdette. New disease-modifying therapies and new challenges for MS. *Current Neurology and Neuroscience Reports*, 12(5):489–491, 2012. ISSN 1528-4042. doi: 10.1007/s11910-012-0295-2. URL http://dx.doi.org/10.1007/s11910-012-0295-2.

[24] Gregory D. Cuny. Neurodegenerative diseases: challenges and opportunities. *Future Medicinal Chemistry*, 4(13):1647–1649, August 2012. ISSN 1756-8919. doi: 10.4155/fmc.12.123. URL http://dx.doi.org/10.4155/fmc.12.123.

[25] Nicholas J. Maragakis and Jeffrey D. Rothstein. Mechanisms of disease: astrocytes in neurodegenerative disease. *Nature Clinical Practice Neurology*, 2(12):679–689, December 2006. ISSN 1745-834X. URL http://dx.doi.org/10.1038/ncpneuro0355.

[26] Jill Stein, Ted Schettler, Ben Rohrer, Maria Valenti, and Nancy Myers. Environmental threats to healthy aging. With a closer look at Alzheimer's and Parkinson's diseases. Technical report, Greater Boston Physicians for Social Responsibility and Science and Envrionmental Health Network, 2008. URL http://www.agehealthy.org/pdf/GBPSRSEHN_HealthyAging1017.pdf.

[27] Adrian Williams. Defining neurodegenerative diseases. *BMJ*, 324(7352):1465–1466, 6 2002. doi: 10.1136/bmj.324.7352.1465.

[28] Richard A. Armstrong. On the 'classification' of neurodegenerative disorders: discrete entities, overlap or continuum? *Folia Neuropathologica*, 50:201–218, 2012.

[29] Anne B. Young. Four decades of neurodegenerative disease research: How far we have come! *The Journal of Neuroscience*, 29(41):12722–12728, 2009. doi: 10.1523/JNEUROSCI.3767-09.2009. URL `http://www.jneurosci.org/content/29/41/12722.short`.

[30] S. L. Miksys and R. F. Tyndale. Neurodegenerative diseases: A growing challenge. *Clinical Pharmacology and Therapeutics*, 88(4):427–430, October 2010. ISSN 0009-9236. URL `http://dx.doi.org/10.1038/clpt.2010.198`.

[31] Erika Pastrana. Neuroscience: A handle on neurodegenerative disease complexity. *Nature Methods*, 9(1):21–21, January 2012. ISSN 1548-7091. URL `http://dx.doi.org/10.1038/nmeth.1847`.

[32] Alexey Kolodkin, Evangelos Simeonidis, Rudi Balling, and Hans Westerhoff. Understanding complexity in neurodegenerative diseases: in silico reconstruction of emergence. *Frontiers in Physiology*, 3(291), 2012. ISSN 1664-042X. doi: 10.3389/fphys.2012.00291. URL `http://www.frontiersin.org/systems_biology/10.3389/fphys.2012.00291/abstract`.

[33] Jorrit J. Hornberg, Frank J. Bruggeman, Hans V. Westerhoff, and Jan Lankelma. Cancer: A systems biology disease. *Biosystems*, 83(23):81–90, 2006. ISSN 0303-2647. doi: 10.1016/j.biosystems.2005.05.014. URL `http://www.sciencedirect.com/science/article/pii/S0303264705001176`.

[34] Andrea Cavalli, Maria Laura Bolognesi, Anna Minarini, Michela Rosini, Vincenzo Tumiatti, Maurizio Recanatini, and Carlo Melchiorre. Multi-target-directed ligands to combat neurodegenerative diseases. *Journal of Medicinal Chemistry*, 51(3):347–372, 2008. doi: 10.1021/jm7009364. URL `http://pubs.acs.org/doi/abs/10.1021/jm7009364`. PMID: 18181565.

[35] Moussa B.H. Youdim and Jerry J. Buccafusco. Multi-functional drugs for various CNS targets in the treatment of neurodegenerative disorders. *Trends in Pharmacological Sciences*, 26(1):27–35, 2005. ISSN 0165-6147. doi: 10.1016/j.tips.2004.11.007. URL `http://www.sciencedirect.com/science/article/pii/S016561470400313X`.

[36] Shan Zhao and Ravi Iyengar. Systems Pharmacology: Network analysis to identify multiscale mechanisms of drug action. *Annual Review of Pharmacology and Toxicology*, 52(1):505–521, 2012. doi: 10.1146/annurev-pharmtox-010611-134520. URL `http://www.annualreviews.org/doi/abs/10.1146/annurev-pharmtox-010611-134520`. PMID: 22235860.

[37] Sang Hong Lee, Julius H. J. van der Werf, Ben J. Hayes, Michael E. Goddard, and Peter M. Visscher. Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS Genetics*, 4(10):e1000231, October 2008. doi: 10.1371/journal.pgen.1000231. URL `http://dx.doi.org/10.1371%2Fjournal.pgen.1000231`.

[38] Mohammad Shahid, Karl N. Kirschner, and Martin Hofmann-Apitius. Pharmacome Tools: A Cytoscape plugin to filter large drug-protein networks for predicting new drug-protein associations. Submitted. *Bioinformatics*, 2013.

[39] Mohammad Shahid, Muhammad Shahzad Cheema, Alexander Klenner, Erfan Younesi, and Martin Hofmann-Apitius. SVM based descriptor selection and classification of neurodegenerative disease drugs for pharmacological modeling. *Molecular Informatics*, 32(3):241–249, 2013. ISSN 1868-1751. doi: 10.1002/minf.201200116. URL `http://dx.doi.org/10.1002/minf.201200116`.

[40] Mohammad Shahid, Martin Hofmann-Apitius, Oliver Waeldrich, and Wolfgang Ziegler. A robust framework for rapid deployment of a virtual screening laboratory. *Studies in Health Technology and Informatics*, 147:212–221, 2009.

[41] Mohammad Shahid, Vinod Kasam, and Martin Hofmann-Apitius. An improved weighted-residue profile based method of using protein-ligand interaction information in increasing hits selection from virtual screening: A study on virtual screening of human GPCR A2A receptor antagonists. *Molecular Informatics*, 29(11):781–791, 2010. ISSN 1868-1751. doi: 10.1002/minf.201000068. URL `http://dx.doi.org/10.1002/minf.201000068`.

[42] Brandon Higgs, Michael Elashoff, Sam Richman, and Beata Barci. An online database for brain disease research. *BMC Genomics*, 7(1):70, 2006. ISSN 1471-2164. doi: 10.1186/1471-2164-7-70. URL `http://www.biomedcentral.com/1471-2164/7/70`.

[43] Sharon X. Xie, Young Baek, Murray Grossman, Steven E. Arnold, Jason Karlawish, Andrew Siderowf, Howard Hurtig, Lauren Elman, Leo McCluskey, Vivianna Van Deerlin, Virginia M.Y. Lee, and John Q. Trojanowski. Building an integrated neurodegenerative disease database at an academic health center. *Alzheimers & Dementia*, 7(4):84–93, July 2011. ISSN 1552-5260. URL `http://linkinghub.elsevier.com/retrieve/pii/S1552526010024520?showall=true`.

[44] Ragul Gowthaman, Nithiyadevi Gowthaman, Madhan Kumar Rajangam, and Kalyanaraman Srinivasan. Database of Neurodegenerative Disorders. *Bioinformatics*, 2:153–154, 2007.

[45] Suhas Vasaikar, Aditya Padhi, Bhyravabhotla Jayaram, and James Gomes. NeuroDNet - an open source platform for constructing and analyzing neurodegenerative disease networks. *BMC Neuroscience*, 14(1):3, 2013. ISSN 1471-2202. doi: 10.1186/1471-2202-14-3. URL `http://www.biomedcentral.com/1471-2202/14/3`.

[46] Elodie Portales-Casamar, Carolyn Ch'ng, Frances Lui, Nicolas St-Georges, Anton Zoubarev, Artemis Lai, Mark Lee, Cathy Kwok, Willie Kwok, Luchia Tseng, and Paul Pavlidis. Neurocarta: aggregating and sharing disease-gene relations for the neurosciences. *BMC Genomics*, 14(1):129, 2013. ISSN 1471-2164. doi: 10.1186/1471-2164-14-129. URL `http://www.biomedcentral.com/1471-2164/14/129`.

[47] Li Gong, Ryan P. Owen, Winston Gor, Russ B. Altman, and Teri E. Klein. *PharmGKB: An Integrated Resource of Pharmacogenomic Data and Knowledge*. John Wiley & Sons, Inc., 2002. ISBN 9780471250951. doi: 10.1002/0471250953.bi1407s23. URL `http://dx.doi.org/10.1002/0471250953.bi1407s23`.

[48] Feng Zhu, Zhe Shi, Chu Qin, Lin Tao, Xin Liu, Feng Xu, Li Zhang, Yang Song, Xianghui Liu, Jingxian Zhang, Bucong Han, Peng Zhang, and Yuzong Chen. Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Research*, 40(1):1128–1136, 2012. doi: 10.1093/nar/gkr797. URL `http://nar.oxfordjournals.org/content/40/D1/D1128.abstract`.

[49] Allan Peter Davis, Benjamin L. King, Susan Mockus, Cynthia G. Murphy, Cynthia Saraceni-Richards, Michael Rosenstein, Thomas Wiegers, and Carolyn J. Mattingly. The Comparative Toxicogenomics Database: update 2011. *Nucleic Acids Research*, 39(suppl 1):1067–1072, 2011. doi: 10.1093/nar/gkq813. URL `http://nar.oxfordjournals.org/content/39/suppl_1/D1067.abstract`.

[50] Lars Bertram, Matthew B. McQueen, Kristina Mullin, Deborah Blacker, and Rudolph E. Tanzi. Systematic meta-analyses of Alzheimer Disease genetic association studies: the AlzGene database. *Nature Genetics*, 39(1): 17–23, January 2007. ISSN 1061-4036. URL `http://dx.doi.org/10.1038/ng1934`.

[51] Satoshi Mizuno, Risa Iijima, Soichi Ogishima, Masataka Kikuchi, Yukiko Matsuoka, Samik Ghosh, Tadashi Miyamoto, Akinori Miyashita, Ryozo Kuwano, and Hiroshi Tanaka. AlzPathway: a comprehensive map of signaling pathways of Alzheimer's Disease. *BMC Systems Biology*, 6(1): 52, 2012. ISSN 1752-0509. doi: 10.1186/1752-0509-6-52. URL `http://www.biomedcentral.com/1752-0509/6/52`.

[52] Christina M. Lill, Johannes T. Roehr, Matthew B. McQueen, Fotini K. Kavvoura, Sachin Bagade, Brit-Maren M. Schjeide, Leif M. Schjeide, Esther Meissner, Ute Zauft, Nicole C. Allen, Tian Liu, Marcel Schilling, Kari J. Anderson, Gary Beecham, Daniela Berg, Joanna M. Biernacka, Alexis Brice, Anita L. DeStefano, Chuong B. Do, Nicholas Eriksson, Stewart A. Factor, Matthew J. Farrer, Tatiana Foroud, Thomas Gasser, Taye Hamza, John A. Hardy, Peter Heutink, Erin M. Hill-Burns, Christine Klein, Jeanne C. Latourelle, Demetrius M. Maraganore, Eden R. Martin, Maria Martinez, Richard H. Myers, Michael A. Nalls, Nathan Pankratz, Haydeh Payami, Wataru Satake, William K. Scott, Manu Sharma, Andrew B. Singleton, Kari Stefansson, Tatsushi Toda, Joyce Y. Tung, Jeffery Vance, Nick W. Wood, Cyrus P. Zabetian, Peter Young, Rudolph E. Tanzi, Muin J. Khoury, Frauke Zipp, Hans Lehrach, John P. A. Ioannidis, Lars Bertram, and The International Parkinson's Disease Genomics Consortium (IPDGC) The Parkinson's Disease GWAS Consortium The Wellcome Trust Case Control Consortium 2 (WTCCC2) 23andMe, The Genetic Epidemiology of Parkinson's Disease

(GEO-PD) Consortium. Comprehensive research synopsis and systematic meta-analyses in Parkinson's Disease genetics: The PDGene Database. *PLoS Genetics*, 8(3):e1002548, March 2012. doi: 10.1371/journal.pgen.1002548. URL http://dx.doi.org/10.1371%2Fjournal.pgen.1002548.

[53] C. M. Lill, J. T. Roehr, M. B. McQueen, F. K. Kavvoura, S. Bagade, B.M. Schjeide, F. Zipp, and Bertram L. The MSGene Database. Online, 2013. URL http://www.msgene.org. Accessed on March, 2013.

[54] Ravi Kiran Kalathur, Miguel Hernandez-Prieto, and Matthias Futschik. Huntington's Disease and its therapeutic target genes: a global functional profile based on the HD Research Crossroads database. *BMC Neurology*, 12(1):47, 2012. ISSN 1471-2377. doi: 10.1186/1471-2377-12-47. URL http://www.biomedcentral.com/1471-2377/12/47.

[55] Olubunmi Abel, John F Powell, Peter M. Andersen, and Ammar Al-Chalabi. ALSoD: A user-friendly online bioinformatics tool for Amyotrophic Lateral Sclerosis genetics. *Human Mutation*, 33(9):1345–1351, 2012. ISSN 1098-1004. doi: 10.1002/humu.22157. URL http://dx.doi.org/10.1002/humu.22157.

[56] Sarma A.R.P. Jagarlapudi and K.V. Radha Kishan. Database systems for knowledge-based discovery. In Edgar Jacoby, editor, *Chemogenomics*, volume 575 of *Methods in Molecular Biology*, pages 159–172. Humana Press, 2009. ISBN 978-1-60761-273-5. doi: 10.1007/978-1-60761-274-2_6. URL http://dx.doi.org/10.1007/978-1-60761-274-2_6.

[57] Pekka Tiikkainen and Lutz Franke. Analysis of commercial and public bioactivity databases. *Journal of Chemical Information and Modeling*, 52(2):319–326, 2012. doi: 10.1021/ci2003126. URL http://pubs.acs.org/doi/abs/10.1021/ci2003126.

[58] Carolyn E. Lipscomb. Medical Subject Headings (MeSH). *Bulletin of the Medical Library Association*, 88:265–266, 2000.

[59] Craig Knox, Vivian Law, Timothy Jewison, Philip Liu, Son Ly, Alex Frolkis, Allison Pon, Kelly Banco, Christine Mak, Vanessa Neveu, Yannick Djoumbou, Roman Eisner, An Chi Guo, and David S. Wishart. Drugbank 3.0: a comprehensive resource for omics research on drugs. *Nucleic Acids Research*, 39(suppl 1):D1035–D1041, 2011. doi: 10.1093/nar/gkq1126. URL http://nar.oxfordjournals.org/content/39/suppl_1/D1035.abstract.

[60] Michael Kuhn, Damian Szklarczyk, Andrea Franceschini, Christian von Mering, Lars Juhl Jensen, and Peer Bork. STITCH 3: zooming in on protein-chemical interactions. *Nucleic Acids Research*, 40(D1):D876–D880, 2012. doi: 10.1093/nar/gkr1011. URL http://nar.oxfordjournals.org/content/40/D1/D876.abstract.

[61] Anna Gaulton, Louisa J. Bellis, A. Patricia Bento, Jon Chambers, Mark Davies, Anne Hersey, Yvonne Light, Shaun McGlinchey, David Michalovich,

Bissan Al-Lazikani, and John P. Overington. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research*, 40:1100–1107, 2012.

[62] Xin Wen Robert N. Jorissen Tiqing Liu, Yuhmei Lin and Michael K. Gilson. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Research*, 35:198–201, 2007.

[63] Bryan L. Roth, Estelle Lopez, Shamil Patel, and Wesley K. Kroeze. The multiplicity of serotonin receptors: Uselessly diverse molecules or an embarrassment of riches? *The Neuroscientist*, 6(4):252–262, 2000. doi: 10.1177/107385840000600408. URL `http://nro.sagepub.com/content/6/4/252.abstract`.

[64] Minoru Kanehisa and Susumu Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1):27–30, 2000. doi: 10.1093/nar/28.1.27. URL `http://nar.oxfordjournals.org/content/28/1/27.abstract`.

[65] Michael Kuhn, Monica Campillos, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. A side effect resource to capture phenotypic effects of drugs. *Molecular Systems Biology*, 6:343, January 2010. URL `http://dx.doi.org/10.1038/msb.2009.98`.

[66] Ruth L. Seal, Susan M. Gordon, Michael J. Lush, Mathew W. Wright, and Elspeth A. Bruford. genenames.org: the HGNC resources in 2011. *Nucleic Acids Research*, 39:514–519, 2011. doi: 10.1093/nar/gkq892.

[67] Consurtium. The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Research*, 38(Database-Issue):142–148, 2010.

[68] Williams G.J. Meyer Jr. E.E. Brice M.D. Rodgers J.R. Kennard O. Shimanouchi T. Tasumi M. Bernstein F.C., Koetzle T.F. The Protein Data Bank: A computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, 112:535, 1977.

[69] Nouri Neamati and Joseph J. Barchi. New paradigms in drug design and discovery. *Current Topics in Medicinal Chemistry*, 2:211–227, 2002.

[70] Si Yuan Pan, Shan Pan, Zhi Ling Yu, Dik Lung Ma, Si Bao Chen, Wang Fun Fong, Yi Fan Han, and Kam Ming Ko. New perspectives on innovative drug discovery: An overview. *Journal of Pharmacy and Pharmaceutical Sciences*, 13:450–471, 2010.

[71] S. Müller and J. Weigelt. Open-access public-private partnerships to enable drug discovery–new approaches. *IDrugs*, 13:175–180, 2010.

[72] Avi Ma'ayan, Sherry L. Jenkins, Joseph Goldfarb, and Ravi Iyengar. Network analysis of FDA approved drugs and their targets. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine*, 74 (1):27–32, 2007. ISSN 1931-7581. doi: 10.1002/msj.20002. URL `http://dx.doi.org/10.1002/msj.20002`.

[73] Aihua Zhang, Hui Sun, Bo Yang, and Xijun Wang. Predicting new molecular targets for rhein using network pharmacology. *BMC Systems Biology*, 6 (1):20, 2012. ISSN 1752-0509. doi: 10.1186/1752-0509-6-20. URL `http://www.biomedcentral.com/1752-0509/6/20`.

[74] Lun Yang, Ke-Jian Wang, Li-Shan Wang, AnilG. Jegga, Sheng-Ying Qin, Guang He, Jian Chen, Yue Xiao, and Lin He. Chemical-protein interactome and its application in off-target identification. *Interdisciplinary Sciences: Computational Life Sciences*, 3:22–30, 2011. ISSN 1913-2751. doi: 10.1007/s12539-011-0051-8. URL `http://dx.doi.org/10.1007/s12539-011-0051-8`.

[75] Andreas Brunschweiger and Jonathan Hall. A decade of the human genome sequence: How does the medicinal chemist benefit? *ChemMedChem*, 7(2): 194–203, 2012. ISSN 1860-7187. doi: 10.1002/cmdc.201100498. URL `http://dx.doi.org/10.1002/cmdc.201100498`.

[76] Andrew L. Hopkins. Network pharmacology: the next paradigm in drug discovery. *Nature Chemical Biology*, 4(11):682–690, November 2008. ISSN 1552-4450. URL `http://dx.doi.org/10.1038/nchembio.118`.

[77] Qihe Xu, Fan Qu, and Olavi Pelkonen. *Network Pharmacology and Traditional Chinese Medicine*. 2012. doi: 10.5772/53868. URL `http://www.intechopen.com/books/alternative-medicine/network-pharmacology-and-traditional-chinese-medicine`.

[78] James T. Metz and Philip J. Hajduk. Rational approaches to targeted polypharmacology: creating and navigating protein-ligand interaction networks. *Current Opinion in Chemical Biology*, 14(4):498–504, 2010. ISSN 1367-5931. doi: 10.1016/j.cbpa.2010.06.166. URL `http://www.sciencedirect.com/science/article/pii/S1367593110000712`.

[79] Zoltan Simon, Agnes Peragovics, Vigh-Smeller Margit, Gabor Csukly, Laszlo Tombor, Zhenhui Yang, Gergely Zahoranszky-Kohlami, Laszlo Vegner, Balazs Jelinek, Peter Hari, Csaba Hetenyi, Istvan Bitter, Pal Czobor, and Andras Malnasi-Csizmadia. Drug effect prediction by polypharmacology-based interaction profiling. *Journal of Chemical Information and Modeling*, 52(1):134–145, 2012. doi: 10.1021/ci2002022. URL `http://pubs.acs.org/doi/abs/10.1021/ci2002022`.

[80] Osama M. Alian, Minjel Shah, and Ramzi M. Mohammad. Network pharmacology: Reining in drug attrition? *Current Drug Discovery Technologies*, December 2012. ISSN 1570-1638. URL `http://pubget.com/paper/23237678`.

[81] Georgios Pavlopoulos, Maria Secrier, Charalampos Moschopoulos, Theodoros Soldatos, Sophia Kossida, Jan Aerts, Reinhard Schneider, and Pantelis Bagos. Using graph theory to analyze biological networks. *BioData Mining*, 4(1):10, 2011. ISSN 1756-0381. doi: 10.1186/1756-0381-4-10. URL `http://www.biodatamining.org/content/4/1/10`.

[82] N. Sukumar and Michael P. Krein. Graphs and networks in chemical and biological informatics: past, present and future. *Future Medicinal Chemistry*, 4(16):2039–2047, 2012. doi: doi:10.4155/fmc.12.128. URL http://www.ingentaconnect.com/content/fs/fmc/2012/00000004/00000016/art00006.

[83] Nagasuma Chandra and Jyothi Padiadpu. Network approaches to drug discovery. *Expert Opinion on Drug Discovery*, 8(1):7–20, 2013. doi: 10.1517/17460441.2013.741119. URL http://informahealthcare.com/doi/abs/10.1517/17460441.2013.741119. PMID: 23140510.

[84] Zhong Wang, Jun Liu, Yanan Yu, Yinying Chen, and Yongyan Wang. Modular pharmacology: the next paradigm in drug discovery. *Expert Opinion on Drug Discovery*, 7(8):667–677, 2012. doi: 10.1517/17460441.2012.692673. URL http://informahealthcare.com/doi/abs/10.1517/17460441.2012.692673. PMID: 22680068.

[85] Jennyfer Bultinck, Sam Lievens, and Jan Tavernier. Protein-protein interactions: Network analysis and applications in drug discovery. *Current Pharmaceutical Design*, 18:4619–4629, 2012.

[86] Asfar S. Azmi. Adopting network pharmacology for cancer drug discovery. *Current Drug Discovery Technologies*, 10(2):95–105, 2012.

[87] Jordi Mestres, Elisabet Gregori-Puigjane, Sergi Valverde, and Ricard V Sole. Data completeness–the achilles heel of drug-target networks. *Nature Biotechnology*, 26(9):983–984, September 2008. ISSN 1087-0156. URL http://dx.doi.org/10.1038/nbt0908-983.

[88] Jordi Mestres, Elisabet Gregori-Puigjane, Sergi Valverde, and Ricard V. Sole. The topology of drug-target interaction networks: implicit dependence on drug properties and target families. *Molecular BioSystems*, 5:1051–1057, 2009. doi: 10.1039/B905821B. URL http://dx.doi.org/10.1039/B905821B.

[89] Aislyn D. W. Boran and Ravi Iyengar. Systems pharmacology. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine*, 77(4):333–344, 2010. ISSN 1931-7581. doi: 10.1002/msj.20191. URL http://dx.doi.org/10.1002/msj.20191.

[90] Aislyn D. W. Boran and Ravi Iyengar. Systems approaches to polypharmacology and drug discovery. *Current Opinion in Drug Discovery & Development*, 13:297–309, 2010.

[91] Ingo Vogt and Jordi Mestres. Drug-target networks. *Molecular Informatics*, 29(1-2):10–14, 2010. ISSN 1868-1751. doi: 10.1002/minf.200900069. URL http://dx.doi.org/10.1002/minf.200900069.

[92] Chemical Computing Group Inc. Molecular operating environment (MOE), 2011. URL http://www.chemcomp.com.

[93] Lei Xie and Philip E. Bourne. Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments. *Proceedings of the National Academy of Sciences*, 105(14):5441–5446, 2008. doi: 10.1073/pnas.0704422105. URL `http://www.pnas.org/content/105/14/5441.abstract`.

[94] Lei Xie, Li Xie, and Philip E. Bourne. A unified statistical model to support local sequence order independent similarity searching for ligand-binding sites and its application to genome-based drug discovery. *Bioinformatics*, 25 (12):i305–i312, 2009. doi: 10.1093/bioinformatics/btp220. URL `http://bioinformatics.oxfordjournals.org/content/25/12/i305.abstract`.

[95] Michael E. Smoot, Keiichiro Ono, Johannes Ruscheinski, Peng-Liang Wang, and Trey Ideker. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, 27(3):431–432, 2011. doi: 10.1093/bioinformatics/btq675. URL `http://bioinformatics.oxfordjournals.org/content/27/3/431.abstract`.

[96] M.E. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 69(5):026113, 2004. ISSN 1539-3755.

[97] Gang Su, Allan Kuchinsky, John H. Morris, David J. States, and Fan Meng. GLay: community structure analysis of biological networks. *Bioinformatics*, 26(24):3135–3137, 2010. doi: 10.1093/bioinformatics/btq596. URL `http://bioinformatics.oxfordjournals.org/content/26/24/3135.abstract`.

[98] Symyx Software. Maccs structural keys., 2005.

[99] Darren R. Flower. On the properties of bit string-based measures of chemical similarity. *Journal of Chemical Information and Computer Sciences*, 38(3): 379–386, 1998. doi: 10.1021/ci970437z. URL `http://pubs.acs.org/doi/abs/10.1021/ci970437z`.

[100] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. Ucsf chimera: A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004. ISSN 1096-987X. doi: 10.1002/jcc.20084. URL `http://dx.doi.org/10.1002/jcc.20084`.

[101] Owen Ozier Nitin S. Baliga Jonathan T. Wang Daniel Ramage Nada Amin Benno Schwikowski Trey Ideker Paul Shannon, Andrew Markiel. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13:2498–2504, 2003.

[102] Christoph Steinbeck, Yongquan Han, Stefan Kuhn, Oliver Horlacher, Edgar Luttmann, and Egon Willighagen. The Chemistry Development Kit (CDK): An open-source java library for chemo- and bioinformatics. *Journal of Chemical Information and Computer Sciences*, 43(2):493–500, 2003. doi: 10.1021/ci025584y. URL `http://pubs.acs.org/doi/abs/10.1021/ci025584y`. PMID: 12653513.

[103] Michael R. Berthold Christian Borgelt, Thorsten Meini. Moss: A program for molecular substructure mining. In *Workshop Open Source Data Mining Software*, OSDM'05, Chicago, IL, New York, 2005. ACM Press.

[104] Jun Dong and Steve Horvath. Understanding network concepts in modules. *BMC Systems Biology*, 1(1):24, 2007. ISSN 1752-0509. doi: 10.1186/1752-0509-1-24. URL `http://www.biomedcentral.com/1752-0509/1/24`.

[105] John Morris, Leonard Apeltsin, Aaron Newman, Jan Baumbach, Tobias Wittkop, Gang Su, Gary Bader, and Thomas Ferrin. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics*, 12(1):436, 2011. ISSN 1471-2105. doi: 10.1186/1471-2105-12-436. URL `http://www.biomedcentral.com/1471-2105/12/436`.

[106] A.M. Johnson and G.M. Maggiora. *Concepts and Applications of Molecular Similarity*. John Wiley & Sons, Inc., 1990.

[107] Menelas N. Pangalos, Lee E. Schechter, and Orest Hurko. Drug development for CNS disorders: strategies for balancing risk and reducing attrition. *Nature Reviews Drug Discovery*, 6(7):521–532, July 2007. ISSN 1474-1776. URL `http://dx.doi.org/10.1038/nrd2094`.

[108] Joacobus P. Petzer, Neal Castagnoli, Michael A. Schwarzschild, Jiang Fan Chen, and Cornelis J. Van der Schyf. Dual-target-directed drugs that block monoamine oxidase B and adenosine A2A receptors for Parkinson's disease. *Neurotherapeutics*, 6:141–151, 2009.

[109] Maria Laura Bolognesi, Andrea Cavalli, and Carlo Melchiorre. Memoquin: A multi-target-directed ligand as an innovative therapeutic opportunity for Alzheimer's disease. *Neurotherapeutics*, 6:152–162, 2009.

[110] Narihiro Toda, Keiko Tago, Shinji Marumoto, Kazuko Takami, Mayuko Ori, Naho Yamada, Kazuo Koyama, Shunji Naruto, Kazumi Abe, Reina Yamazaki, Takao Hara, Atsushi Aoyagi, Yasuyuki Abe, Tsugio Kaneko, and Hiroshi Kogen. A conformational restriction approach to the development of dual inhibitors of acetylcholinesterase and serotonin transporter as potential agents for Alzheimer's disease. *Bioorganic & Medicinal Chemistry*, 11(20):4389–4415, 2003. ISSN 0968-0896. doi: 10.1016/S0968-0896(03)00452-8. URL `http://www.sciencedirect.com/science/article/pii/S0968089603004528`.

[111] Y. Xue, Z. R. Li, C. W. Yap, L. Z. Sun, X. Chen, and Y. Z. Chen. Effect of molecular descriptor feature selection in support vector machine classification of pharmacokinetic and toxicological properties of chemical agents. *Journal of Chemical Information and Computer Sciences*, 44(5):1630–1638, 2004. doi: 10.1021/ci049869h. URL `http://pubs.acs.org/doi/abs/10.1021/ci049869h`.

[112] Hu Li, Chun Wei Yap, Choong Yong Ung, Ying Xue, Zhi Wei Cao, and Yu Zong Chen. Effect of selection of molecular descriptors on the prediction

of blood brain barrier penetrating and nonpenetrating agents by statistical learning methods. *Journal of Chemical Information and Modeling*, 45(5): 1376–1384, 2005. doi: 10.1021/ci050135u. URL http://pubs.acs.org/doi/abs/10.1021/ci050135u. PMID: 16180914.

[113] Ron Kohavi and George H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97:273–324, 1997.

[114] Thomas M. Frimurer, Robert Bywater, Lars Nrum, Leif Nrskov Lauritsen, and Sren Brunak. Improving the odds in discriminating drug-like from non drug-like compounds. *Journal of Chemical Information and Computer Sciences*, 40(6):1315–1324, 2000. doi: 10.1021/ci0003810. URL http://pubs.acs.org/doi/abs/10.1021/ci0003810. PMID: 11128089.

[115] Kailin Tang, Ruixin Zhu, Yixue Li, and Zhiwei Cao. Discrimination of approved drugs from experimental drugs by learning methods. *BMC Bioinformatics*, 12(1):157, 2011. ISSN 1471-2105. doi: 10.1186/1471-2105-12-157. URL http://www.biomedcentral.com/1471-2105/12/157.

[116] Mehdi Ghorbanzad'e and Mohammad Hossein Fatemi. Classification of central nervous system agents by least squares support vector machine based on their structural descriptors: A comparative study. *Chemometrics and Intelligent Laboratory Systems*, 110(1):102–107, 2012. ISSN 0169-7439. doi: 10.1016/j.chemolab.2011.10.003. URL http://www.sciencedirect.com/science/article/pii/S0169743911002073.

[117] I. Guyon, J. Weston, S. Barnhill, and V.Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46 (1/3):389–422, 2002. URL http://www.ingentaconnect.com/content/klu/mach/2002/00000046/F0030001/00380515.

[118] Xin Zhou and David P. Tuck. MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data. *Bioinformatics*, 23 (9):1106–1114, 2007. doi: 10.1093/bioinformatics/btm036. URL http://bioinformatics.oxfordjournals.org/content/23/9/1106.abstract.

[119] Thomas A. Halgren. Merck molecular force field. I. basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry*, 17(5-6):490–519, 1996. ISSN 1096-987X. doi: 10.1002/(SICI)1096-987X(199604)17:5/6⟨490::AID-JCC1⟩3.0.CO;2-P. URL http://dx.doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P.

[120] Kai-Bo Duan, Jagath C. Rajapakse, Haiying Wang, and Francisco Azuaje. Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE Transactions on NanoBioscience*, 4:228–234, 2005.

[121] Mikko J. Vainio, J. Santeri Puranen, and Mark S. Johnson. ShaEP: Molecular overlay based on shape and electrostatic potential. *Journal of Chemical Information and Modeling*, 49(2):492–502, 2009. doi: 10.1021/ci800315d. URL http://pubs.acs.org/doi/abs/10.1021/ci800315d.

[122] M. Rarey and J.S. Dixon. Feature Trees: a new molecular similarity measure based on tree matching. *Journal of Computer Aided Molecular Design*, 12 (5):471–490, 1998.

[123] Joseph L. Durant, Burton A. Leland, Douglas R. Henry, and James G. Nourse. Reoptimization of MDL keys for use in drug discovery. *Journal of Chemical Information and Computer Sciences*, 42(6):1273–1280, 2002. doi: 10.1021/ci010132r. URL `http://pubs.acs.org/doi/abs/10.1021/ci010132r`.

[124] Miika Ahdesmaeki and Korbinian Strimmer. Feature selection in omics prediction problems using cat scores and false nondiscovery rate conrol. *Annals of Applied Statistics*, 1:503–519, 2010.

[125] J. Fan and Y. Fan. High dimensional classification using feature annealed independence rules. *Annals of Statistics*, 36:2605–2636, 2008.

[126] Yinglei Lai. Genome-wide co-expression based prediction of differential expressions. *Bioinformatics*, 24(5):666–673, 2008. doi: 10.1093/bioinformatics/btm507. URL `http://bioinformatics.oxfordjournals.org/content/24/5/666.abstract`.

[127] Verena Zuber and Korbinian Strimmer. Gene ranking and biomarker discovery under correlation. *Bioinformatics*, 25(20):2700–2707, 2009. doi: 10.1093/bioinformatics/btp460. URL `http://bioinformatics.oxfordjournals.org/content/25/20/2700.abstract`.

[128] T.I. Oprea, S.K. Nielsen, O. Ursu, J.J. Yang, O. Taboureau, S.L. Mathias, L. Kouskoumvekaki, L.A. Sklar, and C.G. Bologa. Associating drugs, targets and clinical outcomes into an integrated network affords a new platform for computer-aided drug repurposing. *Molecular Informatics*, 30(2-3):100–111, March 2011. ISSN 1611-020X. URL `http://pubget.com/paper/22287994`.

[129] Patrick M. McNeely, Andrea N. Naranjo, and Anne S. Robinson. Structure-function studies with G protein-coupled receptors as a paradigm for improving drug discovery and development of therapeutics. *Biotechnology Journal*, 7(12):1451–1461, 2012. ISSN 1860-7314. doi: 10.1002/biot.201200076. URL `http://dx.doi.org/10.1002/biot.201200076`.

[130] John A. Salon, David T. Lodowski, and Krzysztof Palczewski. The significance of G protein-coupled receptor crystallography for drug discovery. *Pharmacological Reviews*, 63(4):901–937, 2011. doi: 10.1124/pr.110.003350. URL `http://pharmrev.aspetjournals.org/content/63/4/901.abstract`.

[131] Peter Imming, Christian Sinning, and Achim Meyer. Drugs, their targets and the nature and number of drug targets. *Nature Reviews Drug Discovery*, 5(10):821–834, October 2006. ISSN 1474-1776. URL `http://dx.doi.org/10.1038/nrd2132`.

[132] John P. Overington, Bissan Al-Lazikani, and Andrew L. Hopkins. How many drug targets are there? *Nature Reviews Drug Discovery*, 5(12):993–996, December 2006. ISSN 1474-1776. URL http://dx.doi.org/10.1038/nrd2199.

[133] Thomas Klabunde and Gerhard Hessler. Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem*, 3(10):928–944, 2002. ISSN 1439-7633. doi: 10.1002/1439-7633(20021004)3:10⟨928::AID-CBIC928⟩3.0.CO;2-5. URL http://dx.doi.org/10.1002/1439-7633(20021004)3:10<928::AID-CBIC928>3.0.CO;2-5.

[134] Kenneth Lundstrom. An overview on GPCRs and drug discovery: Structure-based drug design and structural biology on GPCRs. In Wayne R. Leifert, editor, *G Protein-Coupled Receptors in Drug Discovery*, volume 552 of *Methods in Molecular Biology*, pages 51–66. Humana Press, 2009. ISBN 978-1-60327-316-9. doi: 10.1007/978-1-60327-317-6_4. URL http://dx.doi.org/10.1007/978-1-60327-317-6_4.

[135] Dorota Latek, Anna Modzelewska, Bartosz Trzaskowski, Krzysztof Palczewski, and Slawomir Filipek. G protein-coupled receptors – recent advances. *Acta Biochimica Polonica*, 59:515–529, 2012.

[136] Rosamaria Lappano and Marcello Maggiolini. G protein-coupled receptors: novel targets for drug discovery in cancer. *Nature Reviews Drug Discovery*, 10(1):47–60, January 2011. ISSN 1474-1776. URL http://dx.doi.org/10.1038/nrd3320.

[137] Ralf Heilker, Michael Wolff, Christofer S. Tautermann, and Michael Bieler. G-protein-coupled receptor-focused drug discovery using a target class platform approach. *Drug Discovery Today*, 14(5-6):231–240, 2009. ISSN 1359-6446. doi: 10.1016/j.drudis.2008.11.011. URL http://www.sciencedirect.com/science/article/pii/S135964460800408X.

[138] Ru Zhang and Xin Xie. Tools for GPCR drug discovery. *Acta Pharmacologica Sinica*, 33(3):372–384, March 2012. ISSN 1671-4083. URL http://dx.doi.org/10.1038/aps.2011.173.

[139] Wesley K. Kroeze, Douglas J. Sheffler, and Bryan L. Roth. G-protein-coupled receptors at a glance. *Journal of Cell Science*, 116(24):4867–4869, 2003. doi: 10.1242/jcs.00902. URL http://jcs.biologists.org/content/116/24/4867.short.

[140] Edgar Jacoby, Rochdi Bouhelal, Marc Gerspacher, and Klaus Seuwen. The 7-TM G-protein-coupled receptor target family. *ChemMedChem*, 1(8):760–782, 2006. ISSN 1860-7187. doi: 10.1002/cmdc.200600134. URL http://dx.doi.org/10.1002/cmdc.200600134.

[141] P. Michael Conn, Alfredo Ulloa-Aguirre, Joel Ito, and Jo Ann Janovick. G protein-coupled receptor trafficking in health and disease: Lessons learned

to prepare for therapeutic mutant rescue in vivo. *Pharmacological Reviews*, 59(3):225–250, 2007. doi: 10.1124/pr.59.3.2. URL http://pharmrev.aspetjournals.org/content/59/3/225.abstract.

[142] Bronwen Martin, Rakel Lopez Maturana, Randall Brenneman, Tom Walent, MarkP. Mattson, and Stuart Maudsley. Class II G protein-coupled receptors and their ligands in neuronal function and protection. *NeuroMolecular Medicine*, 7:3–36, 2005. ISSN 1535-1084. doi: 10.1385/NMM:7:1-2:003. URL http://dx.doi.org/10.1385/NMM%3A7%3A1-2%3A003.

[143] M. Maccarrone, G. Bernardi, A. Agr, and D. Centonze. Cannabinoid receptor signaling in neurodegenerative diseases: a potential role for membrane fluidity disturbance. *British journal of pharmacology*, 163(7):1379–1390, August 2011. ISSN 0007-1188. doi: 10.1111/j.1476-5381.2011.01277.x. URL http://hdl.handle.net/2108/51211.

[144] Megan J. Dowie, Emma L. Scotter, Emanuela Molinari, and Michelle Glass. The therapeutic potential of G-protein coupled receptors in Huntington's disease. *Pharmacology & Therapeutics*, 128(2):305–323, 2010. ISSN 0163-7258. doi: 10.1016/j.pharmthera.2010.07.008. URL http://www.sciencedirect.com/science/article/pii/S0163725810001518.

[145] Mauro Maccarrone, Natalia Battista, and Diego Centonze. The endocannabinoid pathway in Huntington's disease: A comparison with other neurodegenerative diseases. *Progress in Neurobiology*, 81(5-6):349–379, 2007. ISSN 0301-0082. doi: 10.1016/j.pneurobio.2006.11.006. URL http://www.sciencedirect.com/science/article/pii/S0301008206001419.

[146] Xiaosong Liu and Jian Zhao. GPCR, a rider of Alzheimer's disease. *Frontiers in Biology*, 6(4):282–288, 2011. ISSN 1674-7984. doi: 10.1007/s11515-011-1129-3. URL http://dx.doi.org/10.1007/s11515-011-1129-3.

[147] Tiziana Bisogno and Vincenzo Di Marzo. Cannabinoid receptors and endocannabinoids: Role in neuroinflammatory and neurodegenerative disorders. *CNS & Neurological Disorders - Drug Targets*, 9:564–573, 2010.

[148] Megan C. Chapter, Caitlin M. White, Angela DeRidder, Wayne Chadwick, Bronwen Martin, and Stuart Maudsley. Chemical modification of class II G protein-coupled receptor ligands: Frontiers in the development of peptide analogs as neuroendocrine pharmacological therapies. *Pharmacology & Therapeutics*, 125(1):39–54, 2010. ISSN 0163-7258. doi: 10.1016/j.pharmthera.2009.07.006. URL http://www.sciencedirect.com/science/article/pii/S0163725809001600.

[149] Donald R. Staines. Are Multiple Sclerosis and Amyotrophic Lateral Sclerosis autoimmune disorders of endogenous vasoactive neuropeptides? *Medical Hypotheses*, 70(2):413–418, 2008. ISSN 0306-9877. doi: 10.1016/j.mehy.2007.04.038. URL http://www.sciencedirect.com/science/article/pii/S0306987707003295.

[150] N. Tabet. Acetylcholinesterase inhibitors for Alzheimer's disease: anti-inflammatories in acetylcholine clothing! *Age and Ageing*, 35(4):336–338, 2006. doi: 10.1093/ageing/afl027. URL `http://ageing.oxfordjournals.org/content/35/4/336.abstract`.

[151] John W. Wright and Joseph W. Harding. Importance of the brain angiotensin system in Parkinson's disease. *Parkinson's Disease*, 2012:14, 2012. doi: 10.1155/2012/860923. URL `http://dx.doi.org/10.1155/2012/860923`.

[152] Makoto Naoi and Wakako Maruyama. Functional mechanism of neuroprotection by inhibitors of type B monoamine oxidase in Parkinson's disease. *Expert Reviews Neurotherapeutics*, 9(8):1233–1250, August 2009. ISSN 1473-7175. doi: 10.1586/ern.09.68. URL `http://dx.doi.org/10.1586/ern.09.68`.

[153] Kjell G. Fuxe, Alexander O. Tarakanov, Larisa B. Goncharova, and Luigi F. Agnati. A new road to neuroinflammation in Parkinson's disease? *Brain Research Reviews*, 58(2):453–458, 2008. ISSN 0165-0173. doi: 10.1016/j.brainresrev.2008.04.003. URL `http://www.sciencedirect.com/science/article/pii/S0165017308000325`.

[154] R. Guixa-Gonzalez, A. Bruno, M. Marti-Solano, and J. Selent. Crosstalk within GPCR heteromers in Schizophrenia and Parkinson's disease: Physical or just functional? *Current Medicinal Chemistry*, 19:1119–1134, 2012.

[155] M Bhowmik, R Khanam, and D Vohora. Histamine H3 receptor antagonists in relation to epilepsy and neurodegeneration: a systemic consideration of recent progress and perspectives. *British Journal of Pharmacology*, 167(7):1398–1414, 2012. ISSN 1476-5381. doi: 10.1111/j.1476-5381.2012.02093.x. URL `http://dx.doi.org/10.1111/j.1476-5381.2012.02093.x`.

[156] Laura Albizu, Jose L. Moreno, Javier Gonzalez-Maeso, and Stuart C. Sealfon. Heteromerization of G protein-coupled receptors: Relevance to neurological disorders and neurotherapeutics. *CNS & Neurological Disorders – Drug Targets*, 1:636–650, 2010.

[157] Marie Therese Armentero, Annalisa Pinna, Sergi Ferre, Jos Luis Lanciego, Christa E. Müller, and Rafael Franco. Past, present and future of A2A adenosine receptor antagonists in the therapy of Parkinson's disease. *Pharmacology & Therapeutics*, 132(3):280–299, 2011. ISSN 0163-7258. doi: 10.1016/j.pharmthera.2011.07.004. URL `http://www.sciencedirect.com/science/article/pii/S0163725811001616`.

[158] D. Boison, J.F. Chen, and B. B. Fredholm. Adenosine signaling and function in glial cells. *Cell Death & Differentiation*, 17(7):1071–1082, July 2010. ISSN 1350-9047. URL `http://dx.doi.org/10.1038/cdd.2009.131`.

[159] Alan Grossfield. Recent progress in the study of G protein-coupled receptors with molecular dynamics computer simulations. *Biochimica et Biophysica*

Acta (BBA) - Biomembranes, 1808(7):1868–1878, 2011. ISSN 0005-2736. doi: 10.1016/j.bbamem.2011.03.010. URL http://www.sciencedirect.com/science/article/pii/S0005273611000873.

[160] Anat Levit, Dov Barak, Maik Behrens, Wolfgang Meyerhof, and MashaY. Niv. Homology model-assisted elucidation of binding sites in GPCRs. In Nagarajan Vaidehi and Judith Klein-Seetharaman, editors, *Membrane Protein Structure and Dynamics*, volume 914 of *Methods in Molecular Biology*, pages 179–205. Humana Press, 2012. ISBN 978-1-62703-022-9. doi: 10.1007/978-1-62703-023-6_11. URL http://dx.doi.org/10.1007/978-1-62703-023-6_11.

[161] Mengyuan Zhu and Minyong Li. Revisiting the homology modeling of G-protein coupled receptors: β-1-adrenoceptor as an example. *Molecular BioSystems*, 8:1686–1693, 2012. doi: 10.1039/C2MB05491D. URL http://dx.doi.org/10.1039/C2MB05491D.

[162] Stefano Costanzi. Homology modeling of class A G protein-coupled receptors. In Andrew J. W. Orry and Ruben Abagyan, editors, *Homology Modeling*, volume 857 of *Methods in Molecular Biology*, pages 259–279. Humana Press, 2012. ISBN 978-1-61779-587-9. doi: 10.1007/978-1-61779-588-6_11. URL http://dx.doi.org/10.1007/978-1-61779-588-6_11.

[163] Nathanael Weil. Chemogenomic approaches for the exploration of GPCR space. *Current Topics in Medicinal Chemistry*, 11:1944–1955, 2011.

[164] P. Jimonet and R. Jaeger. Strategies for designing GPCR-focused libraries and screening sets. *Current Opinion in Drug Discovery & Development*, 7(3):325–333, May 2004. URL http://europepmc.org/abstract/MED/15216936.

[165] Konstantin V. Balakin, Sergey E. Tkachenko, Stanley A. Lang, Ilya Okun, Andrey A. Ivashchenko, and Nikolay P. Savchuk. Property-based design of GPCR-targeted library. *Journal of Chemical Information and Computer Sciences*, 42(6):1332–1342, 2002. doi: 10.1021/ci025538y. URL http://pubs.acs.org/doi/abs/10.1021/ci025538y.

[166] Miles Congreve, Christopher Langmead, and Fiona H. Marshall. Chapter 1 - the use of GPCR structures in drug design. In Richard R. Neubig, editor, *Pharmacology of G Protein Coupled Receptors*, volume 62 of *Advances in Pharmacology*, pages 1–36. Academic Press, 2011. doi: 10.1016/B978-0-12-385952-5.00011-7. URL http://www.sciencedirect.com/science/article/pii/B9780123859525000117.

[167] Jonathan S. Mason, Andrea Bortolato, Miles Congreve, and Fiona H. Marshall. New insights from structural biology into the druggability of G protein-coupled receptors. *Trends in Pharmacological Sciences*, 33(5):249–260, May 2012. ISSN 0165-6147. URL http://linkinghub.elsevier.com/retrieve/pii/S0165614712000314.

[168] Stefano Costanzi, Irina G. Tikhonova, T. Kendall Harden, and Kenneth A. Jacobson. Ligand and structure-based methodologies for the prediction of the activity of G protein-coupled receptor ligands. *Journal of Computer-Aided Molecular Design*, 23:747–754, 2009. ISSN 0920-654X. doi: 10.1007/s10822-008-9218-3. URL `http://dx.doi.org/10.1007/s10822-008-9218-3`.

[169] Bas Vroling, Marijn Sanders, Coos Baakman, Annika Borrmann, Stefan Verhoeven, Jan Klomp, Laerte Oliveira, Jacob de Vlieg, and Gert Vriend. GPCRDB: information system for G protein-coupled receptors. *Nucleic Acids Research*, 39(suppl 1):D309–D319, 2011. doi: 10.1093/nar/gkq1009. URL `http://nar.oxfordjournals.org/content/39/suppl_1/D309.abstract`.

[170] Jeroen Kazius, Kerstin Wurdinger, Maarten van Iterson, Joost Kok, Thomas Böck, and Ad P. IJzerman. GPCR NaVa database: natural variants in human G protein-coupled receptors. *Human Mutation*, 29(1):39–44, 2008. ISSN 1098-1004. doi: 10.1002/humu.20638. URL `http://dx.doi.org/10.1002/humu.20638`.

[171] Yasushi Okuno, Jiyoon Yang, Kei Taneishi, Hiroaki Yabuuchi, and Gozoh Tsujimoto. GLIDA: GPCR-ligand database for chemical genomic drug discovery. *Nucleic Acids Research*, 34(suppl 1):D673–D677, 2006. doi: 10.1093/nar/gkj028. URL `http://nar.oxfordjournals.org/content/34/suppl_1/D673.abstract`.

[172] Edgar A. Gatica and Claudio N. Cavasotto. Ligand and decoy sets for docking to G protein-coupled receptors. *Journal of Chemical Information and Modeling*, 52(1):1–6, 2012. doi: 10.1021/ci200412p. URL `http://pubs.acs.org/doi/abs/10.1021/ci200412p`.

[173] Niu Huang, Brian K. Shoichet, and John J. Irwin. Benchmarking sets for molecular docking. *Journal of Medicinal Chemistry*, 49(23):6789–6801, 2006. doi: 10.1021/jm0608356. URL `http://pubs.acs.org/doi/abs/10.1021/jm0608356`.

[174] F. Lauer and Y. Guermeur. MSVMpack: a multi-class support vector machine package. *Journal of Machine Learning Research*, 12:2269–2272, 2011.

[175] S. Ekins, A. Bugrim, L. Brovold, E. Kirillov, Y. Nikolsky, E. Rakhmatulin, S. Sorokina, A. Ryabov, T. Serebryiskaya, A. Melnikov, J. Metz, and T. Nikolskaya. Algorithms for network analysis in systems-ADME/Tox using the MetaCore and MetaDrug platforms. *Xenobiotica*, 36(10-11):877–901, 2006. doi: 10.1080/00498250600861660. URL `http://informahealthcare.com/doi/abs/10.1080/00498250600861660`. PMID: 17118913.

[176] Francesco Iorio, Roberta Bosotti, Emanuela Scacheri, Vincenzo Belcastro, Pratibha Mithbaokar, Rosa Ferriero, Loredana Murino, Roberto Tagliaferri,

Nicola Brunetti-Pierri, Antonella Isacchi, and Diego di Bernardo. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proceedings of the National Academy of Sciences*, 107(33):14621–14626, 2010. doi: 10.1073/pnas.1000138107. URL `http://www.pnas.org/content/107/33/14621.abstract`.

[177] Justin Lamb. The connectivity map: a new tool for biomedical research. *Nature Reviews Cancer*, 7(1):54–60, January 2007. ISSN 1474-175X. URL `http://dx.doi.org/10.1038/nrc2044`.

[178] Justin Lamb, Emily D. Crawford, David Peck, Joshua W. Modell, Irene C. Blat, Matthew J. Wrobel, Jim Lerner, Jean-Philippe Brunet, Aravind Subramanian, Kenneth N. Ross, Michael Reich, Haley Hieronymus, Guo Wei, Scott A. Armstrong, Stephen J. Haggarty, Paul A. Clemons, Ru Wei, Steven A. Carr, Eric S. Lander, and Todd R. Golub. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935, 2006. doi: 10.1126/science.1132939. URL `http://www.sciencemag.org/content/313/5795/1929.abstract`.

[179] T. Langer and R.D. Hoffmann. Virtual screening: An effective tool for lead structure discovery. *Current Pharmaceutical Design*, 7:509–527, 2001.

[180] A. Srinivas Reddy, S. Priyadarshini Pati, P.Praveen Kumar, H.N. Pradeep, and G.Narahari Sastry. Virtual screening in drug discovery - a computational perspective. *Current Protein and Peptide Science*, 8:329–351, 2007.

[181] Gerhard Klebe. Recent developments in structure-based drug design. *Journal of Molecular Medicine*, 78:269–281, 2000. ISSN 0946-2716. doi: 10.1007/s001090000084. URL `http://dx.doi.org/10.1007/s001090000084`.

[182] A.N. Jain. Virtual screening in lead discovery and optimization. *Current Opinion in Drug Discovery & Development*, 7:396–403, 2004.

[183] Douglas B. Kitchen, Helene Decornez, John R. Furr, and Jürgen Bajorath. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery*, 3(11):935–949, November 2004. ISSN 1474-1776. URL `http://dx.doi.org/10.1038/nrd1549`.

[184] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, and C. R. Corbeil. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *British Journal of Pharmacology*, 153 (S1):S7–S26, 2008. ISSN 1476-5381. doi: 10.1038/sj.bjp.0707515. URL `http://dx.doi.org/10.1038/sj.bjp.0707515`.

[185] S.C. Brewerton. The use of protein-ligand interaction fingerprints in docking. *Current Opinion in Drug Discovery and Development*, 11:356–64, 2008.

[186] Hans F. G. Velec, Holger Gohlke, and Gerhard Klebe. DrugScoreCSD: Knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *Journal of Medicinal Chemistry*, 48(20):6296–6303, 2005.

doi: 10.1021/jm050436v. URL http://pubs.acs.org/doi/abs/10.1021/jm050436v.

[187] Patrick Pfeffer, Gerd Neudert, and Gerhard Klebe. DrugScoreFP: profiling protein-ligand interactions using fingerprint simplicity paired with knowledge-based potential fields. *Chemistry Central Journal*, 2(Suppl 1): S16, 2008. ISSN 1752-153X. doi: 10.1186/1752-153X-2-S1-S16. URL http://www.journal.chemistrycentral.com/content/2/S1/S16.

[188] Gilles Marcou and Didier Rognan. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *Journal of Chemical Information and Modeling*, 47(1):195–207, 2007. doi: 10.1021/ci600342e. URL http://pubs.acs.org/doi/abs/10.1021/ci600342e.

[189] Noriaki Okimoto, Noriyuki Futatsugi, Hideyoshi Fuji, Atsushi Suenaga, Gentaro Morimoto, Ryoko Yanai, Yousuke Ohno, Tetsu Narumi, and Makoto Taiji. High-performance drug discovery: Computational screening by combining docking and molecular dynamics simulations. *PLoS Computational Biology*, 5(10):e1000528, October 2009. doi: 10.1371/journal.pcbi.1000528. URL http://dx.doi.org/10.1371%2Fjournal.pcbi.1000528.

[190] Gianluca Degliesposti, Vinod Kasam, Ana DaCosta, Hee-Kyoung Kang, Nahyun Kim, Do-Won Kim, Vincent Breton, Doman Kim, and Giulio Rastelli. Design and discovery of plasmepsinii inhibitors using an automated workflow on large-scale grids. *ChemMedChem*, 4(7):1164–1173, 2009. ISSN 1860-7187. doi: 10.1002/cmdc.200900111. URL http://dx.doi.org/10.1002/cmdc.200900111.

[191] Matthew D. Kelly and Ricardo L. Mancera. Expanded interaction fingerprint method for analyzing ligand binding modes in docking and structure-based drug design. *Journal of Chemical Information and Computer Sciences*, 44 (6):1942–1951, 2004. doi: 10.1021/ci049870g. URL http://pubs.acs.org/doi/abs/10.1021/ci049870g. PMID: 15554663.

[192] Zhan Deng, Claudio Chuaqui, and Juswinder Singh. Structural interaction fingerprint (SIFt): A novel method for analyzing three-dimensional protein-ligand binding interactions. *Journal of Medicinal Chemistry*, 47(2):337–344, 2004. doi: 10.1021/jm030331x. URL http://pubs.acs.org/doi/abs/10.1021/jm030331x. PMID: 14711306.

[193] Josel I. Borrell Jordi Teixido Violeta I. Pelrez-Nueno, Obdulia Rabal. APIF: A new interaction fingerprint based on atom pairs and its application to virtual screening. *Journal of Chemical Information and Modeling*, 49(5): 1245–1260, 2009. doi: 10.1021/ci900043r. URL http://pubs.acs.org/doi/abs/10.1021/ci900043r.

[194] Claudio Chuaqui, Zhan Deng, and Juswinder Singh. Interaction profiles of protein kinase inhibitor complexes and their application to virtual screening. *Journal of Medicinal Chemistry*, 48(1):121–133, 2005.

doi: 10.1021/jm049312t. URL http://pubs.acs.org/doi/abs/10.1021/jm049312t. PMID: 15634006.

[195] Ravi K. Nandigam, Sangtae Kim, Juswinder Singh, and Claudio Chuaqui. Position specific interaction dependent scoring technique for virtual screening based on weighted protein-ligand interaction fingerprint profiles. *Journal of Chemical Information and Modeling*, 49(5):1185–1192, 2009. doi: 10.1021/ci800466n. URL http://pubs.acs.org/doi/abs/10.1021/ci800466n. PMID: 19415918.

[196] Ashutosh Kumar and Mohammad Imran Siddiqi. Virtual screening against mycobacterium tuberculosis dihydrofolate reductase: Suggested workflow for compound prioritization using structure interaction fingerprints. *Journal of Molecular Graphics and Modelling*, 27(4):476 – 488, 2008. ISSN 1093-3263. doi: 10.1016/j.jmgm.2008.08.005. URL http://www.sciencedirect.com/science/article/pii/S1093326308001010.

[197] Steffen Renner, Swetlana Derksen, Sebastian Radestock, and Fabian Morchen. Maximum common binding modes (MCBM): Consensus docking scoring using multiple ligand information and interaction fingerprints. *Journal of Chemical Information and Modeling*, 48(2):319–332, 2008. doi: 10.1021/ci7003626. URL http://pubs.acs.org/doi/abs/10.1021/ci7003626. PMID: 18211051.

[198] Andrew C. Wallace, Roman A. Laskowski, and Janet M. Thornton. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering*, 8(2):127–134, 1995. doi: 10.1093/protein/8.2.127. URL http://peds.oxfordjournals.org/content/8/2/127.abstract.

[199] Roman A. Laskowski. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Research*, 29(1):221–222, 2001. doi: 10.1093/nar/29.1.221. URL http://nar.oxfordjournals.org/content/29/1/221.abstract.

[200] Katrin Stierand and Matthias Rarey. Drawing the PDB: Protein-ligand complexes in two dimensions. *ACS Medicinal Chemistry Letters*, 1(9):540–545, 2010. doi: 10.1021/ml100164p. URL http://pubs.acs.org/doi/abs/10.1021/ml100164p.

[201] Matthias Rarey, Bernd Kramer, Thomas Lengauer, and Gerhard Klebe. A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, 261(3):470 – 489, 1996. ISSN 0022-2836. doi: 10.1006/jmbi.1996.0477. URL http://www.sciencedirect.com/science/article/pii/S0022283696904775.

[202] Garrett M. Morris, David S. Goodsell, Robert S. Halliday, Ruth Huey, William E. Hart, Richard K. Belew, and Arthur J. Olson. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, 19(14):1639–1662, 1998. ISSN 1096-987X. doi: 10.1002/(SICI)1096-987X(19981115)

19:14⟨1639::AID-JCC10⟩3.0.CO;2-B. URL http://dx.doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.

[203] Vinod Kasam, Marc Zimmermann, Astrid Maass, Horst Schwichtenberg, Antje Wolf, Nicolas Jacq, Vincent Breton, and Martin Hofmann-Apitius. Design of new plasmepsin inhibitors: A virtual high throughput screening approach on the egee grid. *Journal of Chemical Information and Modeling*, 47(5):1818–1828, 2007. doi: 10.1021/ci600451t. URL http://pubs.acs.org/doi/abs/10.1021/ci600451t. PMID: 17727268.

[204] Sergio Filipe Sousa, Pedro Alexandrino Fernandes, and Maria Joao Ramos. Protein-ligand docking: Current status and future challenges. *Proteins: Structure, Function, and Bioinformatics*, 65(1):15–26, 2006. ISSN 1097-0134. doi: 10.1002/prot.21082. URL http://dx.doi.org/10.1002/prot.21082.

[205] M.F. Sanner. Python: a programming language for software integration and development. *Journal of Molecular Graphics and Modelling*, 17:57–61, 1999.

[206] Veli-Pekka Jaakola, Mark T. Griffith, Michael A. Hanson, Vadim Cherezov, Ellen Y. T. Chien, J. Robert Lane, Adriaan P. IJzerman, and Raymond C. Stevens. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science*, 322(5905):1211–1217, 2008. doi: 10.1126/science.1164772. URL http://www.sciencemag.org/content/322/5905/1211.abstract.

[207] Marie-Christine Galas Patrizia Popoli Serge N Schiffmann David Blum, Raphael Hourez. Adenosine receptors and Huntington's disease: implications for pathogenesis and therapeutics. *The Lancet Neurology*, 2(6):366–374, 2003. ISSN 1474-4422. doi: 10.1016/S1474-4422(03)00411-3. URL http://www.sciencedirect.com/science/article/pii/S1474442203004113.

[208] Giovannella Strappaghetti, Stefano Corsano, Roberta Barbaro, Gino Giannaccini, and Laura Betti. Structure-activity relationships in a series of 8-substituted xanthines as A1-adenosine receptor antagonists. *Bioorganic & Medicinal Chemistry*, 9(3):575–583, 2001. ISSN 0968-0896. doi: 10.1016/S0968-0896(00)00271-6. URL http://www.sciencedirect.com/science/article/pii/S0968089600002716.

[209] Christa E. Müller, Mark Thorand, Ramatullah Qurishi, Martina Diekmann, Kenneth A. Jacobson, William L. Padgett, and John W. Daly. Imidazo[2,1-i]purin-5-ones and related tricyclic water-soluble purine derivatives: Potent A2A- and A3-adenosine receptor antagonists. *Journal of Medicinal Chemistry*, 45(16):3440–3450, 2002. doi: 10.1021/jm011093d. URL http://pubs.acs.org/doi/abs/10.1021/jm011093d. PMID: 12139454.

[210] Mantri Monica, de Graaf Olivier, van Veldhoven Jacobus, Goeblyoes Anika, von Frijtag Drabbe, Kuenzel Jacobien K., Mulder-Krieger Thea, Link Regina, de Vries Henk, Beukers Margot W., Brussee Johannes, and IJzerman Adriaan P. 2-amino-6-furan-2-yl-4-substituted nicotinonitriles as A2A

adenosine receptor antagonists. *Journal of Medicinal Chemistry*, 51(15): 4449–4455, 2008. doi: 10.1021/jm701594y. URL http://pubs.acs.org/doi/abs/10.1021/jm701594y. PMID: 18637670.

[211] Marcus H. Holschbach, Dirk Bier, Stefan Stuesgen, Walter Wutz, Wiebke Sihver, Heinz H. Coenen, and Ray A. Olsson. Synthesis andevaluation of7-amino-2-(2(3)-furyl)-5-phenylethylamino-oxazolo[5,4-d]pyrimidines aspotential A2A adenosine receptor antagonists forpositron emission tomography (PET). *European Journal of Medicinal Chemistry*, 41(1):7–15, 2006. ISSN 0223-5234. doi: 10.1016/j.ejmech.2005.07.018. URL http://www.sciencedirect.com/science/article/pii/S022352340500245X.

[212] John J. Irwin and Brian K. Shoichet. ZINC: A free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*, 45(1):177–182, 2005. doi: 10.1021/ci049714+. URL http://pubs.acs.org/doi/abs/10.1021/ci049714%2B. PMID: 15667143.

[213] Uta Lessel, Bernd Wellenzohn, Markus Lilienthal, and Holger Claussen. Searching fragment spaces with Feature Trees. *Journal of Chemical Information and Modeling*, 49(2):270–279, 2009. doi: 10.1021/ci800272a. URL http://pubs.acs.org/doi/abs/10.1021/ci800272a.

[214] Marcel L. Verdonk, Valerio Berdini, Michael J. Hartshorn, Wijnand T. M. Mooij, Christopher W. Murray, Richard D. Taylor, and Paul Watson. Virtual screening using protein-ligand docking: Avoiding artificial enrichment. *Journal of Chemical Information and Computer Sciences*, 44(3):793–806, 2004. doi: 10.1021/ci034289q. URL http://pubs.acs.org/doi/abs/10.1021/ci034289q. PMID: 15154744.

[215] OpenEye. Openeye: Scientific software. http://www.eyesopen.com, 2010. cited: 2010.

[216] S. M. Poucher, J.R. Keddie, P.Singh, S.M. Stoggall, P.W. Caulkett, G. Jones, and M.G. Coll. The in vitro pharmacology of ZM241385, a potent, non-xanthine A2A selective adenosine receptor antagonist. *British journal of pharmacology*, 115:1096–1102, 1995.

[217] A. Dearing. Computer-aided molecular modelling: Research study or research tool? *Journal of Computer-Aided Molecular Design*, 2:179–189, 1988. ISSN 0920-654X. doi: 10.1007/BF01531992. URL http://dx.doi.org/10.1007/BF01531992.

[218] Sandhya Kortagere, Markus Lill, and John Kerrigan. Role of computational methods in pharmaceutical sciences. In *Computational Toxicology*, volume 929 of *Methods in Molecular Biology*, pages 21–48. Humana Press, 2012. ISBN 978-1-62703-049-6. doi: 10.1007/978-1-62703-050-2_3. URL http://dx.doi.org/10.1007/978-1-62703-050-2_3.

[219] Si-sheng Ou-Yang, Jun-yan Lu, Xiang-qian Kong, Zhong-jie Liang, Cheng Luo, and Hualiang Jiang. Computational drug discovery. *Acta Pharmacologica Sinica*, 33(9):1131–1140, September 2012. ISSN 1671-4083. URL http://dx.doi.org/10.1038/aps.2012.109.

[220] S. Ekins, J. Mestres, and B. Testa. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *British Journal of Pharmacology*, 152(1):9–20, 2007. ISSN 1476-5381. doi: 10.1038/sj.bjp.0707305. URL http://dx.doi.org/10.1038/sj.bjp.0707305.

[221] Campbell McInnes. Virtual screening strategies in drug discovery. *Current Opinion in Chemical Biology*, 11(5):494–502, 2007. ISSN 1367-5931. doi: 10.1016/j.cbpa.2007.08.033. URL http://www.sciencedirect.com/science/article/pii/S1367593107001172.

[222] Ortega S. Santiago, Cara L. C. Lopez, and Maria K. Salvador. In silico pharmacology for a multidisciplinary drug discovery process. *Drug Metabolism and Drug Interactions*, 27:199–207, 2012. ISSN 21910162. URL http://www.degruyter.com/view/j/dmdi.2012.27.issue-4/dmdi-2012-0021/dmdi-2012-0021.xml.

[223] Diane Joseph-McCarthy. Computational approaches to structure-based ligand design. *Pharmacology & Therapeutics*, 84(2):179–191, 1999. ISSN 0163-7258. doi: 10.1016/S0163-7258(99)00031-5. URL http://www.sciencedirect.com/science/article/pii/S0163725899000315.

[224] J.P. Hughes, S. Rees, S.B. Kalindjian, and K.L. Philpott. Principles of early drug discovery. *British Journal of Pharmacology*, 162(6):1239–1249, 2011. ISSN 1476-5381. doi: 10.1111/j.1476-5381.2010.01127.x. URL http://dx.doi.org/10.1111/j.1476-5381.2010.01127.x.

[225] Michael D. Rawlins. Cutting the cost of drug development? *Nature Reviews Drug Discovery*, 3(4):360–364, April 2004. ISSN 1474-1776. URL http://dx.doi.org/10.1038/nrd1347.

[226] Bruno O. Villoutreix, Richard Eudes, and Maria A. Miteva. Structure-based virtual ligand screening: Recent success stories. *Combinatorial Chemistry and High Throughput Screening*, 12:1000–1016, 2009.

[227] Hans Matter and Christoph Sotriffer. *Applications and Success Stories in Virtual Screening*, pages 319–358. Wiley-VCH Verlag GmbH & Co. KGaA, 2011. ISBN 9783527633326. doi: 10.1002/9783527633326.ch12. URL http://dx.doi.org/10.1002/9783527633326.ch12.

[228] Markus H.J. Seifert and Martin Lang. Essential factors for successful virtual screening. *Mini Reviews in Medicinal Chemistry*, 8:63–72, 2008.

[229] Steven W. Muchmore, Jeremy J. Edmunds, Kent D. Stewart, and Philip J. Hajduk. Cheminformatic tools for medicinal chemists. *Journal of Medicinal Chemistry*, 53(13):4830–4841, 2010. doi: 10.1021/jm100164z. URL http://pubs.acs.org/doi/abs/10.1021/jm100164z. PMID: 20329801.

[230] Subha Kalyaanamoorthy and Yi-Ping Phoebe Chen. Structure-based drug design to augment hit discovery. *Drug Discovery Today*, 16(1718):831 – 839, 2011. ISSN 1359-6446. doi: 10.1016/j.drudis.2011.07.006. URL http://www.sciencedirect.com/science/article/pii/S1359644611002194.

[231] Gisbert Schneider. Virtual screening: an endless staircase? *Nature Reviews Drug Discovery*, 9(4):273–276, April 2010. ISSN 1474-1776. URL http://dx.doi.org/10.1038/nrd3139.

[232] Youssef L. Bennani. Drug discovery in the next decade: innovation needed asap. *Drug Discovery Today*, 17, Supplement(0):S31–S44, 2012. ISSN 1359-6446. doi: 10.1016/j.drudis.2011.12.007. URL http://www.sciencedirect.com/science/article/pii/S135964461100434X.

[233] Peter Ripphausen, Britta Nisius, Lisa Peltason, and Jurgen Bajorath. Quo vadis, virtual screening? a comprehensive survey of prospective applications. *Journal of Medicinal Chemistry*, 53(24):8461–8467, 2010. doi: 10.1021/jm101020z. URL http://pubs.acs.org/doi/abs/10.1021/jm101020z.

[234] W. Patrick Walters, Matthew T. Stahl, and Mark A. Murcko. Virtual screening: an overview. *Drug Discovery Today*, 3(4):160–178, 1998. ISSN 1359-6446. doi: 10.1016/S1359-6446(97)01163-X. URL http://www.sciencedirect.com/science/article/pii/S135964469701163X.

[235] Lorenz M. Mayr and Dejan Bojanic. Novel trends in high-throughput screening. *Current Opinion in Pharmacology*, 9(5):580–588, 2009. ISSN 1471-4892. doi: 10.1016/j.coph.2009.08.004. URL http://www.sciencedirect.com/science/article/pii/S1471489209001283.

[236] Yusuf Tanrikulu, Björn Kräger, and Ewgenij Proschak. The holistic integration of virtual screening in drug discovery. *Drug Discovery Today*, 2013. ISSN 1359-6446. doi: 10.1016/j.drudis.2013.01.007. URL http://www.sciencedirect.com/science/article/pii/S135964461300024X.

[237] Pedro J. Ballester, Martina Mangold, Nigel I. Howard, Richard L. Marchese Robinson, Chris Abell, Jochen Blumberger, and John B. O. Mitchell. Hierarchical virtual screening for the discovery of new molecular scaffolds in antibacterial hit identification. *Journal of The Royal Society Interface*, 2012. doi: 10.1098/rsif.2012.0569. URL http://rsif.royalsocietypublishing.org/content/early/2012/08/23/rsif.2012.0569.abstract.

[238] Cristian G. Bologa and Tudor I. Oprea. Compound collection preparation for virtual screening. In *Bioinformatics and Drug Discovery*, volume 910 of *Methods in Molecular Biology*, pages 125–143. Humana Press, 2012. ISBN 978-1-61779-964-8. doi: 10.1007/978-1-61779-965-5_7. URL http://dx.doi.org/10.1007/978-1-61779-965-5_7.

[239] Tiejun Cheng, Qingliang Li, Zhigang Zhou, Yanli Wang, and StephenH. Bryant. Structure-based virtual screening for drug discovery: a problem-centric review. *The AAPS Journal*, 14:133–141, 2012.

doi: 10.1208/s12248-012-9322-0. URL http://dx.doi.org/10.1208/s12248-012-9322-0.

[240] Peter Ripphausen, Britta Nisius, and Jürgen Bajorath. State-of-the-art in ligand-based virtual screening. *Drug Discovery Today*, 16(9-10):372–376, 2011. ISSN 1359-6446. doi: 10.1016/j.drudis.2011.02.011. URL http://www.sciencedirect.com/science/article/pii/S1359644611000626.

[241] Tudor I. Oprea and Hans Matter. Integrating virtual screening in lead discovery. *Current Opinion in Chemical Biology*, 8(4):349–358, 2004. ISSN 1367-5931. doi: 10.1016/j.cbpa.2004.06.008. URL http://www.sciencedirect.com/science/article/pii/S1367593104000845.

[242] Soma Mandal, Mee'nal Moudgil, and Sanat K. Mandal. Rational drug design. *European Journal of Pharmacology*, 625(1-3):90–100, 2009. ISSN 0014-2999. doi: 10.1016/j.ejphar.2009.06.065. URL http://www.sciencedirect.com/science/article/pii/S0014299909008784.

[243] Martin Schneider. A rational approach to maximize success rate in target discovery. *Archiv der Pharmazie*, 337(12):625–633, 2004. ISSN 1521-4184. doi: 10.1002/ardp.200400913. URL http://dx.doi.org/10.1002/ardp.200400913.

[244] Frank Sams-Dodd. Target-based drug discovery: is something wrong? *Drug Discovery Today*, 10(2):139 – 147, 2005. ISSN 1359-6446. doi: 10.1016/S1359-6446(04)03316-1. URL http://www.sciencedirect.com/science/article/pii/S1359644604033161.

[245] Ahmet Sacan, Sean Ekins, and Sandhya Kortagere. Applications and limitations of in silico models in drug discovery. In *Bioinformatics and Drug Discovery*, volume 910 of *Methods in Molecular Biology*, pages 87–124. Humana Press, 2012. ISBN 978-1-61779-964-8. doi: 10.1007/978-1-61779-965-5_6. URL http://dx.doi.org/10.1007/978-1-61779-965-5_6.

[246] Dagmar Stumpfe, Peter Ripphausen, and Jürgen Bajorath. Virtual compound screening in drug discovery. *Future Medicinal Chemistry*, 4(5):593–602, March 2012. ISSN 1756-8919. doi: 10.4155/fmc.12.19. URL http://dx.doi.org/10.4155/fmc.12.19.

[247] Jeff Blaney. A very short history of structure-based design: how did we get here and where do we need to go? *Journal of Computer-Aided Molecular Design*, 26:13–14, 2012. ISSN 0920-654X. doi: 10.1007/s10822-011-9518-x. URL http://dx.doi.org/10.1007/s10822-011-9518-x.

[248] I.M. Kapetanovic. Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. *Chemico-Biological Interactions*, 171(2):165–176, 2008. ISSN 0009-2797. doi: 10.1016/j.cbi.2006.12.006. URL http://www.sciencedirect.com/science/article/pii/S0009279706003541.

[249] Kathrin Heikamp and Jürgen Bajorath. The future of virtual compound screening. *Chemical Biology & Drug Design*, 81(1):33–40, 2013. ISSN 1747-0285. doi: 10.1111/cbdd.12054. URL http://dx.doi.org/10.1111/cbdd.12054.

[250] Joanna Schaffhausen. Advances in structure-based drug design. *Trends in Pharmacological Sciences*, 33(5):223, 2012. ISSN 0165-6147. doi: 10.1016/j.tips.2012.03.011. URL http://www.sciencedirect.com/science/article/pii/S0165614712000430.

[251] Pter Csermely, Vilmos goston, and Sndor Pongor. The efficiency of multi-target drugs: the network approach might help drug design. *Trends in Pharmacological Sciences*, 26(4):178 – 182, 2005. ISSN 0165-6147. doi: 10.1016/j.tips.2005.02.007. URL http://www.sciencedirect.com/science/article/pii/S0165614705000556.

[252] Arthur M Doweyko and Lidia M Doweyko. What is next for small-molecule drug discovery? *Future Medicinal Chemistry*, 1(6):1029–1036, September 2009. ISSN 1756-8919. doi: 10.4155/fmc.09.72. URL http://dx.doi.org/10.4155/fmc.09.72.

[253] Zachary A. Knight, Henry Lin, and Kevan M. Shokat. Targeting the cancer kinome through polypharmacology. *Nature Reviews Cancer*, 10(2):130–137, February 2010. ISSN 1474-175X. URL http://dx.doi.org/10.1038/nrc2787.

[254] Alejandro Merino, Agnieszka K. Bronowska, David B. Jackson, and Dolores J. Cahill. Drug profiling: knowing where it hits. *Drug Discovery Today*, 15(17-18):749–756, 2010. ISSN 1359-6446. doi: 10.1016/j.drudis.2010.06.006. URL http://www.sciencedirect.com/science/article/pii/S1359644610001984.

[255] Jin-Jian Lu, Wei Pan, Yuan-Jia Hu, and Yi-Tao Wang. Multi-target drugs: The trend of drug research and development. *PLoS ONE*, 7(6):e40262, 06 2012. doi: 10.1371/journal.pone.0040262. URL http://dx.doi.org/10.1371%2Fjournal.pone.0040262.

[256] M.L. Bolognesi, A. Minarini, V. Tumiatti, and C. Melchiorre. Lipoic acid, a lead structure for multi-target-directed drugs for neurodegeneration. *Mini Reviews in Medicinal Chemistry*, 6(11):1269–1274, November 2006. URL http://europepmc.org/abstract/MED/17100639.

[257] Andrew L. Hopkins, Jonathan S. Mason, and John P. Overington. Can we rationally design promiscuous drugs? *Current Opinion in Structural Biology*, 16(1):127–136, 2006. ISSN 0959-440X. doi: 10.1016/j.sbi.2006.01.013. URL http://www.sciencedirect.com/science/article/pii/S0959440X06000157.

[258] Joanne Kotz. Drug discovery: Polypharmacology on the fly. *Nature Chemical Biology*, 8(8):679–679, August 2012. ISSN 1552-4450. URL http://dx.doi.org/10.1038/nchembio.1036.

[259] Ye Hu and Jürgen Bajorath. Compound promiscuity: what can we learn from current data? *Drug Discovery Today*, 2013. ISSN 1359-6446. doi: http://dx.doi.org/10.1016/j.drudis.2013.03.002. URL `http://www.sciencedirect.com/science/article/pii/S1359644613000706`.

[260] Jeremy Besnard, Gian Filippo Ruda, Vincent Setola, Keren Abecassis, Ramona M. Rodriguiz, Xi-Ping Huang, Suzanne Norval, Maria F. Sassano, Antony I. Shin, Lauren A. Webster, Frederick R. C. Simeons, Laste Stojanovski, Annik Prat, Nabil G. Seidah, Daniel B. Constam, G. Richard Bickerton, Kevin D. Read, William C. Wetsel, Ian H. Gilbert, Bryan L. Roth, and Andrew L. Hopkins. Automated design of ligands to polypharmacological profiles. *Nature*, 492(7428):215–220, December 2012. ISSN 0028-0836. URL `http://dx.doi.org/10.1038/nature11691`.

[261] Maria Laura Bolognesi. Polypharmacology in a single drug: multitarget drugs. *Current Medicinal Chemistry*, 20(13):1639–1645, 2013.

[262] Ye Hu and Jürgen Bajorath. Polypharmacology directed compound data mining: Identification of promiscuous chemotypes with different activity profiles and comparison to approved drugs. *Journal of Chemical Information and Modeling*, 50(12):2112–2118, 2010. doi: 10.1021/ci1003637. URL `http://pubs.acs.org/doi/abs/10.1021/ci1003637`.

[263] Beth Apsel, Jimmy A. Blair, Beatriz Gonzalez, Tamim M. Nazif, Morri E. Feldman, Brian Aizenstein, Randy Hoffman, Roger L. Williams, Kevan M. Shokat, and Zachary A. Knight. Targeted polypharmacology: discovery of dual inhibitors of tyrosine and phosphoinositide kinases. *Nature Chemical Biology*, 4(11):691–699, November 2008. ISSN 1552-4450. URL `http://dx.doi.org/10.1038/nchembio.117`.

[264] Thomas Mendgen, Christian Steuer, and Christian D. Klein. Privileged scaffolds or promiscuous binders: A comparative study on rhodanines and related heterocycles in medicinal chemistry. *Journal of Medicinal Chemistry*, 55(2):743–753, 2012. doi: 10.1021/jm201243p. URL `http://pubs.acs.org/doi/abs/10.1021/jm201243p`.

[265] Howard J. Feldman and Paul Labute. Pocket similarity: Are α carbons enough? *Journal of Chemical Information and Modeling*, 50(8):1466–1475, 2010. doi: 10.1021/ci100210c. URL `http://pubs.acs.org/doi/abs/10.1021/ci100210c`.

[266] Francesca Milletti and Anna Vulpetti. Predicting polypharmacology by binding site similarity: From kinases to the protein universe. *Journal of Chemical Information and Modeling*, 50(8):1418–1431, 2010. doi: 10.1021/ci1001263. URL `http://pubs.acs.org/doi/abs/10.1021/ci1001263`.

[267] Irene Nobeli, Angelo D. Favia, and Janet M. Thornton. Protein promiscuity and its implications for biotechnology. *Nature Biotechnology*, 27(2):157–167, February 2009. ISSN 1087-0156. URL `http://dx.doi.org/10.1038/nbt1519`.

[268] Grant R. Zimmermann, Joseph Lehr, and Curtis T. Keith. Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discovery Today*, 12(12):34 – 42, 2007. ISSN 1359-6446. doi: 10.1016/j.drudis.2006.11.008. URL http://www.sciencedirect.com/science/article/pii/S1359644606004697.

[269] C. Lindsay DeVane and Charles B. Nemeroff. An evaluation of risperidone drug interactions. *Journal of Clinical Psychopharmacology*, 4:408–416, 2001.

[270] Ian Gilron and Mitchell B. Max. Combination pharmacotherapy for neuropathic pain: current evidence and future directions. *Expert Reviews of Neurotherapeutics*, 5(6):823–830, November 2005. ISSN 1473-7175. doi: 10.1586/14737175.5.6.823. URL http://dx.doi.org/10.1586/14737175.5.6.823.

[271] Xing-Ming Zhao, Murat Iskar, Georg Zeller, Michael Kuhn, Vera van Noort, and Peer Bork. Prediction of drug combinations by integrating molecular and pharmacological data. *PLoS Computational Biology*, 7(12):e1002323, December 2011. doi: 10.1371/journal.pcbi.1002323. URL http://dx.doi.org/10.1371%2Fjournal.pcbi.1002323.

[272] Hernn Alonso, Andrey A. Bliznyuk, and Jill E. Gready. Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews*, 26(5):531–568, 2006. ISSN 1098-1128. doi: 10.1002/med.20067. URL http://dx.doi.org/10.1002/med.20067.

[273] Jacob Durrant and J. Andrew McCammon. Molecular dynamics simulations and drug discovery. *BMC Biology*, 9(1):71, 2011. ISSN 1741-7007. doi: 10.1186/1741-7007-9-71. URL http://www.biomedcentral.com/1741-7007/9/71.

[274] Xiaolin Li, Li Ye, Xiaoxiang Wang, Xinzhou Wang, Hongling Liu, Xiangping Qian, Yongliang Zhu, and Hongxia Yu. Molecular docking, molecular dynamics simulation, and structure-based 3D-QSAR studies on estrogenic activity of hydroxylated polychlorinated biphenyls. *Science of The Total Environment*, 441:230–238, 2012. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2012.08.072. URL http://www.sciencedirect.com/science/article/pii/S0048969712011631.

[275] Jianling Liu, Hong Zhang, Zhengtao Xiao, Fangfang Wang, Xia Wang, and Yonghua Wang. the $\beta 5$ subunit of human 20S proteasome. *International Journal of Molecular Sciences*, 12(3):1807–1835, 2011. ISSN 1422-0067. doi: 10.3390/ijms12031807. URL http://www.mdpi.com/1422-0067/12/3/1807.

[276] Gerald M. Maggiora. Is there a future for computational chemistry in drug research? *Journal of Computer-Aided Molecular Design*, 26:87–90, 2012. ISSN 0920-654X. doi: 10.1007/s10822-011-9493-2. URL http://dx.doi.org/10.1007/s10822-011-9493-2.

[277] Marcin Nowosielski, Marcin Hoffmann, Aneta Kuron, Malgorzata Korycka-Machala, and Jaroslaw Dziadek. The MM2QM tool for combining docking, molecular dynamics, molecular mechanics, and quantum mechanics. *Journal of Computational Chemistry*, 2012. ISSN 1096-987X. doi: 10.1002/jcc.23192. URL `http://dx.doi.org/10.1002/jcc.23192`.

[278] Shivani Agarwal, Deepak Dugar, and Shiladitya Sengupta. Ranking chemical structures for drug discovery: A new machine learning approach. *Journal of Chemical Information and Modeling*, 50(5):716–731, 2010. doi: 10.1021/ci9003865. URL `http://pubs.acs.org/doi/abs/10.1021/ci9003865`. PMID: 20387860.

[279] Xue-Gang Yang, Wei Lv, Yu-Zong Chen, and Ying Xue. In silico prediction and screening of gamma-secretase inhibitors by molecular descriptors and machine learning methods. *Journal of Computational Chemistry*, 31(6):1249–1258, 2010. ISSN 1096-987X. doi: 10.1002/jcc.21411. URL `http://dx.doi.org/10.1002/jcc.21411`.

[280] Jacob D. Durrant and J. Andrew McCammon. NNScore 2.0: A neural-network receptor-ligand scoring function. *Journal of Chemical Information and Modeling*, 51(11):2897–2903, 2011. doi: 10.1021/ci2003889. URL `http://pubs.acs.org/doi/abs/10.1021/ci2003889`.

[281] Giulio Rastelli, Alberto Del Rio, Gianluca Degliesposti, and Miriam Sgobba. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *Journal of Computational Chemistry*, 31(4):797–810, 2010. ISSN 1096-987X. doi: 10.1002/jcc.21372. URL `http://dx.doi.org/10.1002/jcc.21372`.

[282] Miriam Sgobba, Fabiana Caporuscio, Andrew Anighoro, Corinne Portioli, and Giulio Rastelli. Application of a post-docking procedure based on MM-PBSA and MM-GBSA on single and multiple protein conformations. *European Journal of Medicinal Chemistry*, 58:431–440, 2012. ISSN 0223-5234. doi: 10.1016/j.ejmech.2012.10.024. URL `http://www.sciencedirect.com/science/article/pii/S0223523412006289`.

[283] Giulio Rastelli, Gianluca Degliesposti, Alberto Del Rio, and Miriam Sgobba. Binding estimation after refinement, a new automated procedure for the refinement and rescoring of docked ligands in virtual screening. *Chemical Biology & Drug Design*, 73(3):283–286, 2009. ISSN 1747-0285. doi: 10.1111/j.1747-0285.2009.00780.x. URL `http://dx.doi.org/10.1111/j.1747-0285.2009.00780.x`.

[284] Atsushi Suenaga, Noriaki Okimoto, Yoshinori Hirano, and Kazuhiko Fukui. An efficient computational method for calculating ligand binding affinities. *PLoS ONE*, 7(8):e42846, 08 2012. doi: 10.1371/journal.pone.0042846. URL `http://dx.doi.org/10.1371%2Fjournal.pone.0042846`.

[285] Tingjun Hou, Junmei Wang, Youyong Li, and Wei Wang. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. the accuracy

of binding free energy calculations based on molecular dynamics simulations. *Journal of Chemical Information and Modeling*, 51(1):69–82, 2011. doi: 10.1021/ci100275a. URL http://pubs.acs.org/doi/abs/10.1021/ci100275a.

[286] David W. Borhani and David E. Shaw. The future of molecular dynamics simulations in drug discovery. *Journal of Computer-Aided Molecular Design*, 26:15–26, 2012. ISSN 0920-654X. doi: 10.1007/s10822-011-9517-y. URL http://dx.doi.org/10.1007/s10822-011-9517-y.

[287] Matthew J. Harvey and Gianni De Fabritiis. High-throughput molecular dynamics: the powerful new tool for drug discovery. *Drug Discovery Today*, 17(1920):1059 – 1062, 2012. ISSN 1359-6446. doi: 10.1016/j.drudis.2012.03.017. URL http://www.sciencedirect.com/science/article/pii/S1359644612001213.

[288] BioSolveIT. Leadit. http://www.biosolveit.de/leadit, 2012. URL http://www.biosolveit.de/leadit.

[289] D.A. Case, T.A. Darden, T.E. Cheatham, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossvary, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, , and P.A. Kollman. Amber 12. University of California, San Francisco., 2012. URL http://ambermd.org.

[290] Andreas W. Goetz, Mark J. Williamson, Dong Xu, Duncan Poole Scott Le Grand, and Ross C. Walker. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born. *Journal of Chemical Theory and Computation*, 8(5):1542–1555, 2012. doi: 10.1021/ct200909j. URL http://pubs.acs.org/doi/abs/10.1021/ct200909j.

[291] Bill R. Miller, T. Dwight McGee, Jason M. Swails, Nadine Homeyer, Holger Gohlke, and Adrian E. Roitberg. MMPBSA.py: An efficient program for end-state free energy calculations. *Journal of Chemical Theory and Computation*, 8(9):3314–3321, 2012. doi: 10.1021/ct300418h. URL http://pubs.acs.org/doi/abs/10.1021/ct300418h.

[292] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004. ISSN 1096-987X. doi: 10.1002/jcc.20035. URL http://dx.doi.org/10.1002/jcc.20035.

[293] Eric J. Sorin and Vijay S. Pande. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophysical Journal*, 88(4):2472 – 2493, 2005. ISSN 0006-3495. doi: http://dx.doi.org/10.1529/biophysj.104.

051938. URL http://www.sciencedirect.com/science/article/pii/S000634950573304X.

[294] Jeffry D. Madura Roger W. Impey Michael L.Klein William L. Jorgensen, Jayaraman Chandrasekhar. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79:926–936, 1983.

[295] Asinex, June 2012. URL www.asinex.com/. Accessed June, 2012.

[296] David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010. doi: 10.1021/ci100050t. URL http://pubs.acs.org/doi/abs/10.1021/ci100050t. PMID: 20426451.

[297] Kathrin Heikamp and Jürgen Bajorath. Large-scale similarity search profiling of ChEMBL compound data sets. *Journal of Chemical Information and Modeling*, 51(8):1831–1839, 2011. doi: 10.1021/ci200199u. URL http://pubs.acs.org/doi/abs/10.1021/ci200199u.

[298] Adria Cereto-Massague, Laura Guasch, Cristina Valls, Miquel Mulero, Gerard Pujadas, and Santiago Garcia-Vallve. DecoyFinder: an easy-to-use python GUI application for building target-specific decoy sets. *Bioinformatics*, 28(12):1661–1662, 2012. doi: 10.1093/bioinformatics/bts249. URL http://bioinformatics.oxfordjournals.org/content/28/12/1661.abstract.

[299] Vadim Cherezov, Daniel M. Rosenbaum, Michael A. Hanson, Sren G. F. Rasmussen, Foon Sun Thian, Tong Sun Kobilka, Hee-Jung Choi, Peter Kuhn, William I. Weis, Brian K. Kobilka, and Raymond C. Stevens. High-resolution crystal structure of an engineered human β2-Adrenergic G protein-coupled receptor. *Science*, 318(5854):1258–1265, 2007. doi: 10.1126/science.1150577. URL http://www.sciencemag.org/content/318/5854/1258.abstract.

[300] Ellen Y. T. Chien, Wei Liu, Qiang Zhao, Vsevolod Katritch, Gye Won Han, Michael A. Hanson, Lei Shi, Amy Hauck Newman, Jonathan A. Javitch, Vadim Cherezov, and Raymond C. Stevens. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science*, 330(6007):1091–1095, 2010. doi: 10.1126/science.1197410. URL http://www.sciencemag.org/content/330/6007/1091.abstract.

[301] Tatsuro Shimamura, Mitsunori Shiroishi, Simone Weyand, Hirokazu Tsujimoto, Graeme Winter, Vsevolod Katritch, Ruben Abagyan, Vadim Cherezov, Wei Liu, Gye Won Han, Takuya Kobayashi, Raymond C. Stevens, and So Iwata. Structure of the human histamine H1 receptor complex with doxepin. *Nature*, 475(7354):65–70, July 2011. ISSN 0028-0836. URL http://dx.doi.org/10.1038/nature10236.

[302] Beili Wu, Ellen Y. T. Chien, Clifford D. Mol, Gustavo Fenalti, Wei Liu, Vsevolod Katritch, Ruben Abagyan, Alexei Brooun, Peter Wells, F. Christopher Bi, Damon J. Hamel, Peter Kuhn, Tracy M. Handel, Vadim Cherezov,

and Raymond C. Stevens. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science*, 330(6007):1066–1071, 2010. doi: 10.1126/science.1194396. URL `http://www.sciencemag.org/content/330/6007/1066.abstract`.

[303] Claudia Binda, Frantisek Hublek, Min Li, Yaacov Herzig, Jeffrey Sterling, Dale E. Edmondson, and Andrea Mattevi. Crystal structures of monoamine oxidase B in complex with four inhibitors of the n-propargylaminoindan class. *Journal of Medicinal Chemistry*, 47(7):1767–1774, 2004. doi: 10.1021/jm031087c. URL `http://pubs.acs.org/doi/abs/10.1021/jm031087c`. PMID: 15027868.

[304] Huanchen Wang, Ming-Sheng Peng, Yi Chen, Jie Geng, Howard Robinson, Miles D. Houslay, Jiwen Cai, and Hengming Ke. Structures of the four subfamilies of phosphodiesterase-4 provide insight into the selectivity of their inhibitors. *Biochemical Journal*, 408(2):193–201, 2007. doi: 10.1042/BJ20070970. URL `http://www.biochemj.org/bj/408/bj4080193.htm`.

[305] Hemant Kumar Srivastava and G. Narahari Sastry. Molecular dynamics investigation on a series of HIV protease inhibitors: Assessing the performance of MM-PBSA and MM-GBSA approaches. *Journal of Chemical Information and Modeling*, 52(11):3088–3098, 2012. doi: 10.1021/ci300385h. URL `http://pubs.acs.org/doi/abs/10.1021/ci300385h`.

[306] Tara Mirzadegan, Gil Benko, Slawomir Filipek, and Krzysztof Palczewski. Sequence analyses of G-protein-coupled receptors: Similarities to rhodopsin. *Biochemistry*, 42(10):2759–2767, 2003. doi: 10.1021/bi027224+. URL `http://pubs.acs.org/doi/abs/10.1021/bi027224%2B`. PMID: 12627940.

[307] Henry Lin, Maria F. Sassano, Bryan L. Roth, and Brian K. Shoichet. A pharmacological organization of G protein-coupled receptors. *Nature Methods*, 10(2):140–146, February 2013. ISSN 1548-7091. URL `http://dx.doi.org/10.1038/nmeth.2324`.

[308] D.M. Richards, R.N. Brogden, R.C. Heel, T.M. Speight, and G.S. Avery. Oxatomide. *Drugs*, 27(3):210–231, 1984. ISSN 0012-6667. doi: 10.2165/00003495-198427030-00003. URL `http://dx.doi.org/10.2165/00003495-198427030-00003`.

[309] D.A Walsh, S.K. Frazyshen, and J.M. Janni. Synthesis and antiallergy activity of 4-(diarylhydroxymethyl)-1-[3-(aryloxy)propyl]piperidines and structurally related compounds. *Journal of Medicinal Chemistry*, 32:105–118, 1989.

[310] Judith Varady, Xihan Wu, Xueliang Fang, Ji Min, Zengjian Hu, Beth Levant, and Shaomeng Wang. Molecular modeling of the three-dimensional structure of Dopamine 3 (D3) subtype receptor: Discovery of novel and potent D3 ligands through a hybrid pharmacophore- and structure-based

database searching approach. *Journal of Medicinal Chemistry*, 46(21):4377–4392, 2003. doi: 10.1021/jm030085p. URL `http://pubs.acs.org/doi/abs/10.1021/jm030085p`. PMID: 14521403.

[311] NIH Molecular Libraries Screening Center Network. Inhibitors of T-Type Calcium Channel, 2010. URL `http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=489005`.